

(19) 世界知的所有権機関  
国際事務局

553 431

(43) 国際公開日  
2004 年 11 月 4 日 (04.11.2004)

PCT

(10) 国際公開番号  
WO 2004/095374 A1

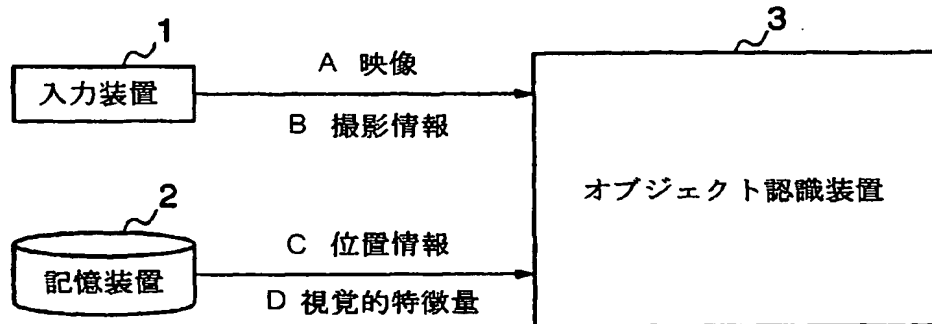
- (51) 国際特許分類<sup>7</sup>: G06T 7/00  
 (21) 国際出願番号: PCT/JP2004/005372  
 (22) 国際出願日: 2004 年 4 月 15 日 (15.04.2004)  
 (25) 国際出願の言語: 日本語  
 (26) 国際公開の言語: 日本語  
 (30) 優先権データ:  
 特願2003-115437 2003 年 4 月 21 日 (21.04.2003) JP  
 (71) 出願人 (米国を除く全ての指定国について): 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目 7 番 1 号 Tokyo (JP).  
 (72) 発明者; および  
 (75) 発明者/出願人 (米国についてのみ): 高橋 祐介 (TAKAHASHI, Yusuke) [JP/JP]; 〒1088001 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内 Tokyo (JP). 平田 恭

- 二 (HIRATA, Kyoji) [JP/JP]; 〒1088001 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内 Tokyo (JP).  
 (74) 代理人: 宮崎 昭夫, 外 (MIYAZAKI, Teruo et al.); 〒1070052 東京都港区赤坂 1 丁目 9 番 2 0 号 第 1 6 興和ビル 8 階 Tokyo (JP).  
 (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.  
 (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL,

[続葉有]

(54) Title: VIDEO OBJECT RECOGNITION DEVICE AND RECOGNITION METHOD, VIDEO ANNOTATION GIVING DEVICE AND GIVING METHOD, AND PROGRAM

(54) 発明の名称: 映像オブジェクト認識装置および認識方法、映像アノテーション付与装置および付与方法ならびにプログラム



- 1...INPUT DEVICE  
 2...STORAGE DEVICE  
 A...VIDEO  
 B...IMAGE PICKUP INFORMATION  
 C...POSITION INFORMATION  
 D...VISUAL FEATURE AMOUNT  
 3...OBJECT RECOGNITION DEVICE

(57) Abstract: Visual feature information as information on a visual feature of an object digitalized is correlated in advance, when stored, with addition information as information added to the object. Partial video data which is video data of a part of video region is extracted. Visual feature information on the extracted partial video data is created and the visual feature information on the partial video data is compared to the visual feature information on the object stored, thereby calculating the similarity. According to the similarity calculated, the object contained in the video data is identified and annotation composed of the addition information of the identified object is superimposed on the video so as to be displayed by a display device.

[続葉有]

WO 2004/095374 A1



SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

(57) 要約: オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とをあらかじめ関連付けて記憶する。そして、映像の一部分の領域の映像データである部分映像データを抽出する。抽出した部分映像データの視覚的特徴情報を生成し、部分映像データの視覚的特徴情報と、記憶しているオブジェクトの視覚的特徴情報とを比較して類似度を算出する。算出した類似度に基づいて映像データに含まれるオブジェクトを特定し、特定したオブジェクトの付加情報によって構成されるアノテーションを映像に重畳して表示装置に表示させる。

## 明細書

映像オブジェクト認識装置および認識方法、映像アノテーション付与装置および付与方法ならびにプログラム

## 技術分野

本発明は、映像の被写体と、被写体の情報とを対応付ける映像認識装置、映像アノテーション付与装置および車両案内システムに関する。

## 背景技術

近年、GPS (global positioning system) を用いて現在位置の情報を利用者に提供するナビゲーションシステムが普及し、自動車等にナビゲーションシステムが搭載されている。これらのナビゲーションシステムには、現在位置に基づいた周辺の情報を利用者に提供する機能があり、ナビゲーションシステムの利用者端末は、提供された情報をモニタ等を介してCG (Computer graphics) で描かれたデジタル地図に表示する。

現在位置の周辺の情報を利用者に提供する他の方法には、強化現実技術 (Augmented Reality) を用いて、周辺の情報を実映像の画面に重畳して表示させる方法がある。

下記の非特許文献1には、実映像の撮影を行う撮影位置と、撮影方向と、映像の被写体であるオブジェクト (建物などのひとまとまりの物体) の緯度および経度の位置情報とを用いて、撮影方向に撮影位置から所定の距離内に存在するオブジェクトの情報 (オブジェクト名) を、実映像に重畳して表示するシステムについて記載されている。

(非特許文献1)

寺田智裕、外2名、「拡張現実感を用いた車載型アノテーションシステムの構築」、信学技報、社団法人電気通信情報学会、2002年2月、CQ2001-103、MVE2001-136、p. 55-60

しかし、非特許文献 1 に記載のシステムは、オブジェクトが映像に写っているか否かに関わらず、撮影方向に撮影位置から所定の距離内に存在するオブジェクトの情報を表示する。そのため、木の陰になる等の理由でオブジェクトが実際には映像に写っていないにも関わらず、オブジェクトの情報を表示してしまうという問題がある。

### 発明の開示

本発明は、オブジェクトが実際には映像に写っていない場合にはオブジェクトの情報を表示しないようにすることができる映像認識装置、映像アノテーション付与装置および車両案内システムを提供することを目的とする。

本発明の映像オブジェクト認識装置は、映像データと、撮影空間を決定しうる情報である撮影情報とを入力する入力手段と、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶する記憶手段と、入力された映像データによる映像に含まれるオブジェクトを認識するオブジェクト認識手段とを備え、オブジェクト認識手段は、撮影情報に基づいて撮影空間を推定する撮影空間推定手段と、撮影空間と、記憶手段に記憶されているオブジェクトの位置情報が示す位置とのマッチングを行うマッチング手段と、入力された映像データから、映像データによる映像の一部分の領域の映像データまたは映像全体の映像データである部分映像データを抽出する部分映像抽出手段と、部分映像データの視覚的特徴情報を生成する視覚的特徴情報設定手段と、部分映像データの視覚的特徴情報と、記憶手段が記憶しているオブジェクトの視覚的特徴情報とを比較して類似度を算出する類似度算出手段と、マッチング手段のマッチングの結果と、算出した類似度とに基づいて、入力された映像データによる映像中にオブジェクトが存在しているか否かを判定する判定手段とを備える。

本発明の映像アノテーション付与装置は、映像データと、撮影空間を決定しうる情報である撮影情報とを入力する入力手段と、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶する

記憶手段と、入力された映像データによる映像に含まれるオブジェクトと、付加情報とを対応付けるオブジェクト認識手段とを備え、オブジェクト認識手段は、撮影情報に基づいて撮影空間を推定する撮影空間推定手段と、撮影空間と、記憶手段に記憶されているオブジェクトの位置情報が示す位置とのマッチングを行うマッチング手段と、入力された映像データから、映像データによる映像の一部分の領域の映像データまたは映像全体の映像データである部分映像データを抽出する部分映像抽出手段と、部分映像データの視覚的特徴情報を生成する視覚的特徴情報設定手段と、部分映像データの視覚的特徴情報と、記憶手段が記憶しているオブジェクトの視覚的特徴情報とを比較して類似度を算出する類似度算出手段と、マッチング手段のマッチングの結果と、算出した類似度とに基づいて、入力された映像データによる映像に含まれるオブジェクトを特定し、特定したオブジェクトと、記憶手段が記憶している付加情報とを対応付ける判定手段とを備える。

また、オブジェクト認識手段は、撮影空間と、記憶手段が記憶しているオブジェクトの位置情報とに基づいてオブジェクトが映像に含まれている確率である存在確率を算出する存在確率算出手段を含んでもよく、判定手段は、算出した存在確率と類似度とに基づいて映像に含まれるオブジェクトを特定し、特定したオブジェクトと記憶手段が記憶している付加情報とを対応付けてもよい。そのような構成によれば、撮影空間に存在するオブジェクトであっても、映像に写っている確率が低いオブジェクトに対しては、付加情報との対応付けを行わない。そのため、アノテーションと映像とを重畳して表示させた場合に、映像に映っていないオブジェクトのアノテーションが表示されることを防ぐことができる。

また、部分映像抽出手段は、記憶手段が記憶しているオブジェクトの位置情報に基づいて、映像におけるオブジェクトが位置する範囲を特定してもよく、特定した範囲から部分映像データを抽出してもよい。そのような構成によれば、部分映像データを抽出する範囲を限定することができ、部分映像抽出手段のデータ処理量を削減させることができる。

また、オブジェクト認識手段は、撮影空間と位置情報とに基づいて撮影空間に存在するオブジェクトである候補オブジェクトの抽出を行う候補オブジェクト検索手段を含んでもよく、類似度算出手段は、部分映像データの視覚的特徴情報と、

記憶手段が記憶している候補オブジェクトの視覚的特徴情報とを比較して類似度を算出してもよい。そのような構成によれば、部分映像データの視覚的特徴情報と比較して類似度を算出するオブジェクトの数を減らすことができ、類似度算出手段のデータ処理量を削減させることができる。

また、部分映像抽出手段は、記憶手段が記憶している候補オブジェクトの位置情報に基づいて、映像におけるオブジェクトが位置する範囲を特定してもよく、特定した範囲の映像から部分映像データを抽出してもよい。そのような構成によれば、部分映像データを抽出する範囲を限定することができ、部分映像抽出手段のデータ処理量を削減させることができる。

また、映像を表示する表示手段と、表示手段に、映像に含まれるオブジェクトに対応付けられた付加情報を表示させる位置を指定して付加情報を映像に重畳して表示させる表示位置決定手段とを備えてもよい。そのような構成によれば、映像と付加情報であるアノテーションとを対応させて表示させることができる。

また、付加情報と映像に含まれるオブジェクトとを対応させて記憶するアノテーション結果記憶手段を備えてもよい。そのような構成によれば、記憶した付加情報と映像データに含まれるオブジェクトとの対応に基づいて、映像データの検索を行うことができる。

また、部分映像抽出手段は、抽出する部分映像データによる映像の領域の形状と大きさを任意に変更可能な機能を有してもよい。そのような構成によれば、映像データに含まれるオブジェクトの形状に関わらず、オブジェクトを含む部分映像データを抽出することができる。

また、部分映像抽出手段は、輝度情報と、色情報と、形状の情報と、模様の情報と、大きさの情報とのうちの1つの条件、または複数を組み合わせた条件に合致する映像の領域の部分映像データを抽出してもよい。そのような構成によれば、記憶手段が記憶するオブジェクトの視覚的特徴情報に応じて、部分映像データを抽出することができる。

また、部分映像抽出手段は、複数の情報を組み合わせた条件に合致する映像の領域の部分映像データを抽出する場合、マッチング手段のマッチングの結果と、記憶手段が記憶しているオブジェクトの視覚的特徴情報とに基づいて、各条件に

対する重み付けを行って部分映像データを抽出してもよい。そのような構成によれば、記憶手段が記憶しているオブジェクトの視覚的な特徴に応じて、部分映像データを抽出することができる。

また、記憶手段が記憶するオブジェクトの視覚的特徴情報は、オブジェクトに類似する視覚的な特徴を有する映像であるテンプレート映像であってもよい。そのような構成によれば、オブジェクトの視覚的特徴情報を記憶手段に記憶させる際の作業量を削減することができる。

また、記憶手段が記憶するオブジェクトの視覚的特徴情報は、色情報、形状の情報、模様の情報および大きさの情報のうち1つ、または複数の情報によって構成されてもよく、視覚的特徴情報設定手段が生成する部分映像データの視覚的特徴情報は、色情報と、形状の情報と、模様の情報と、大きさの情報とのうち1つの情報、または複数の情報によって構成されてもよい。そのような構成によれば、視覚的特徴情報設定手段は定量的に類似度を算出することができる。

また、記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの形状に類似した形状の円錐、円柱、または立方体等の立体幾何学の立体形状を用いて近似した立体形状の頂点のうちの1つの頂点、または中心点、もしくは重心点のいずれかの位置を特定する情報であってもよい。そのような構成によれば、位置情報のデータ量を削減することができる。

また、記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの形状を多角形の面を用いて近似した三次元形状の頂点のうちの少なくとも1つの頂点の位置を特定する情報であってもよい。そのような構成によれば、位置情報のデータ量を削減することができる。

また、記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの頂点のうち、最も標高が高い頂点の位置を特定する情報であってもよい。そのような構成によれば、位置情報のデータ量を削減することができる。

また、記憶手段が記憶するオブジェクトの位置情報は、緯度と経度と標高とによってオブジェクトの位置を特定する情報であってもよい。そのような構成によれば、GPSを用いて位置情報を取得した位置のデータを用いることができる。

また、記憶手段は、オブジェクトの付加情報と、視覚的特徴情報とを、複数のオブジェクトの集合であるオブジェクト集合の付加情報と、視覚的特徴情報とに対して階層的に記憶してもよく、判定手段は、記憶手段が記憶しているオブジェクト集合の視覚的特徴情報に基づいて、撮影された複数のオブジェクトがオブジェクト集合であるか否かの判定を行い、撮影された複数のオブジェクトがオブジェクト集合であると判定すると、オブジェクト集合の付加情報とオブジェクト集合とを対応付けてもよい。そのような構成によれば、オブジェクト集合にアノテーションを対応付けることができる。

また、撮影情報は、撮影した日付と時刻とを特定する情報である撮影日時情報を含んでもよく、記憶手段は、撮影した日付と、時刻とに応じた視覚的特徴情報を記憶してもよく、類似度算出手段は、部分映像データの視覚的特徴情報と、撮影日時情報が特定する日付と、時刻とに応じた視覚的特徴情報とを比較して類似度を算出してもよい。そのような構成によれば、例えば一のオブジェクトに対して昼間の時刻の視覚的特徴情報は明るい色を示す情報とし、夜間の時刻の視覚的特徴情報は暗い色を示す情報とすることができる。また、例えば山等の日付（例えば、季節。）によって色が変化するオブジェクトに対して、冬は白色、春と夏とは緑色、秋は赤色や黄色等を示す視覚的特徴情報にする等、一のオブジェクトの視覚的特徴情報を日付に応じて変化させることができる。そのため、オブジェクト認識手段は、撮影した日付と時刻とによるオブジェクトの視覚的な特徴の変化に応じて、オブジェクトを認識することができる。

また、部分映像抽出手段は、入力された映像データから領域を分割し、分割された領域を部分映像データとしてもよい。

また、部分映像抽出手段は、分割された領域の組み合わせにより部分映像データとしてもよい。

また、部分映像抽出手段は、分割領域の組み合わせの評価を階層的に行うことで部分映像データを生成してもよい。

また、部分映像抽出手段は、分割領域の組み合わせ階層評価として、同一領域数の組み合わせから類似度の高い特定数のみ以後の組み合わせに用いてもよい。



また、記憶手段が記憶するオブジェクトの視覚的特徴情報として、一方向または、複数方向からの全体像または一部分のオブジェクトにおける視覚的情報を複数保持してもよい。

本発明の車両案内システムは、車両に搭載され、GPSを用いて自己の位置を表示装置が表示する地図上に示す車両案内システムであって、上述の映像アノテーション付与装置を備える。

本発明の映像オブジェクト認識方法は、映像データと、撮影空間を決定しうる情報である撮影情報とを入力し、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶し、撮影情報に基づいて撮影空間を推定し、撮影空間と、記憶されているオブジェクトの位置情報が示す位置とのマッチングを行い、入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出し、部分映像データの視覚的特徴情報を生成し、部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出し、マッチングの結果と、算出した類似度とに基づいてオブジェクトが撮影されているか否かを判定する。

本発明の映像アノテーション付与方法は、映像データと、撮影空間を決定しうる情報である撮影情報とを入力し、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶し、撮影情報に基づいて撮影空間を推定し、撮影空間と、記憶されているオブジェクトの位置情報が示す位置とのマッチングを行い、入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出し、部分映像データの視覚的特徴情報を生成し、部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出し、マッチングの結果と、算出した類似度とに基づいて映像データに含まれるオブジェクトを特定し、特定したオブジェクトと記憶されている付加情報によって構成されるアノテーションとを対応付ける。

本発明の映像オブジェクト認識プログラムは、映像データに、記憶されているオブジェクトが被写体として含まれているか否かを判定する映像オブジェクト認識装置に搭載される映像オブジェクト認識プログラムであって、コンピュータに、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶装置に記憶させる処理と、撮影空間を決定しうる情報である撮影情報に基づいて撮影空間を推定する処理と、撮影空間と、記憶装置が記憶しているオブジェクトの位置情報が示す位置とのマッチングを行う処理と、入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出する処理と、部分映像データの視覚的特徴情報を生成する処理と、部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出する処理と、マッチングの結果と、算出した類似度とに基づいてオブジェクトが撮影されているか否かを判定する処理とを実行させる。

本発明の映像アノテーション付与プログラムは、オブジェクトと、記憶されているオブジェクトの情報とを対応付ける映像アノテーション付与装置に搭載される映像アノテーション付与プログラムであって、コンピュータに、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶装置に記憶させる処理と、撮影空間を決定しうる情報である撮影情報に基づいて撮影空間を推定する処理と、撮影空間と、記憶装置が記憶しているオブジェクトの位置情報が示す位置とのマッチングを行う処理と、入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出する処理と、部分映像データの視覚的特徴情報を生成する処理と、部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出する処理と、マッチングの結果と、算出した類似度とに基づいて映像データに含まれるオブジェクトを特定し、特定したオブジェクトと記憶されている付加情報によって構成されるアノテーションとを対応付ける処理とを実行させる。

従って、本発明によれば、オブジェクトの視覚的特徴情報と、部分映像データの視覚的特徴情報とを比較することによって、オブジェクトが映像に映っているか否かの判定を行うことができ、オブジェクトが映像に映っていると判定した場合に、オブジェクトと付加情報とを対応付ける。よって、オブジェクトが実際には映像に写っていない場合にはオブジェクトの情報を表示しないようにすることができる。

また、本発明による車両案内システムは、本発明による映像アノテーション付与装置が備える機能を有するため、オブジェクトが実際には映像に写っていない場合にはオブジェクトの情報を表示しない車両案内システムを実現できる。

#### 図面の簡単な説明

図1は、本発明の実施の形態の一構成例を示すブロック図である。

図2は、意味情報と、位置情報と、視覚的特徴量との例を示す図である。

図3は、本発明の第1の実施の形態の一構成例を示すブロック図である。

図4は、本発明の第1の実施の形態における視覚的特徴量比較手段の一構成例を示すブロック図である。

図5は、本発明の第1の実施の形態の動作を説明するフローチャートである。

図6は、側面から撮影空間を表した図である。

図7は、情報から撮影空間を表した図である。

図8は、撮影空間を説明する図である。

図9は、撮影位置と、候補オブジェクトの代表点との関係の一例を示す図である。

図10は、テンプレート映像の例を示す図である。

図11は、映像全体をラスタスキャンした場合を説明する図である。

図12は、撮影空間の右側に候補オブジェクトが位置する場合の説明図である。

図13は、映像の右側半分をラスタスキャンした場合を説明する図である。

図14は、領域に分割する映像の一例を説明する図である。

図15は、領域の分割と、結合とを説明する図である。

図16は、領域分割した映像の一例を説明する図である。

図 1 7 は、領域の数を減少させた場合の映像の一例を説明する図である。

図 1 8 は、部分映像と候補オブジェクトとを比較する一例を説明する図である。

図 1 9 は、地理オブジェクトの集合の一例を示す図である。

図 2 0 は、階層化された意味情報を説明する図である。

図 2 1 は、地理オブジェクトの一例を正面から見た図と、斜めから見た図とである。

図 2 2 は、分割した地理オブジェクトの一例を示す図である。

図 2 3 は、本発明の第 2 の実施の形態の一構成例を示すブロック図である。

図 2 4 は、本発明の第 2 の実施の形態における視覚的特徴量比較手段の一構成例を示すブロック図である。

図 2 5 は、本発明の第 3 の実施の形態の一構成例を示すブロック図である。

図 2 6 は、アノテーションを付与した映像の一例を示す図である。

図 2 7 は、本発明の第 4 の実施の形態の一構成例を示すブロック図である。

図 2 8 は、本発明の第 5 の実施の形態における視覚的特徴量比較手段の一構成例を示すブロック図である。

図 2 9 は、本発明の第 5 の実施の形態の動作を説明するフローチャートである。

図 3 0 は、階層的組み合わせ評価方法を説明する図である。

### 発明を実施するための最良の形態

#### (発明の第 1 の実施の形態)

図 1 に、本発明の第 1 の実施の形態を説明するブロック図を示し、この実施の形態の構成について説明する。本発明の第 1 の実施の形態は、映像および映像を撮影した状況の情報である撮影情報が入力される入力装置 1 と、映像の被写体であるオブジェクトの情報を記憶する記憶装置 2 と、撮影情報とオブジェクトの情報とを比較して、オブジェクトが映像に撮影されているか否かを判定するオブジェクト認識装置 3 とを含む。

入力装置 1 は、例えば CCD デジタルカメラ装置や、ビデオカメラ装置等の映像を撮影する映像入力機器（図示せず）が出力する映像を入力する。入力装置 1 は、映像入力機器が出力する撮影時の画角の情報である画角情報を入力する。映

像入力機器のズーム率を変更できない場合には、あらかじめレンズの焦点距離およびCCD素子のサイズとから画角を算出して画角情報を生成しておく。ズーム操作によって映像入力機器のズーム率を変更できる場合には、ズーム操作量と画角との関係の情報をあらかじめ取得し、撮影時のズーム操作量に応じた画角を算出して画角情報を生成する。

また、入力装置1は、入力装置1の経度と緯度と標高とを測定するGPS装置と接続され、撮影位置と、移動速度とを特定する。ここで、入力装置1と加速度センサ装置と接続され、加速度センサ装置の出力を時間積分して移動速度を特定してもよい。また、入力装置1は、磁気コンパス装置、電子コンパス装置、またはジャイロコンパス装置と接続され、映像入力機器が撮影する方向を特定する。

ここで、映像入力機器がパン・チルト機能を有する場合、パン・チルト操作量を測定して、磁気コンパス装置、電子コンパス装置、またはジャイロコンパス装置が特定した撮影方向の補正を行う。また、映像入力機器が撮影中に移動した場合、入力装置1は、GPS装置の軌跡データや、電子コンパス装置、ジャイロコンパス装置、加速度センサ装置等の出力を用いて、移動方向を特定する。撮影情報は、画角情報、撮影した日付と時刻とを特定する情報である撮影日時情報、撮影位置の情報、撮影方向の情報、移動方向の情報、および移動速度の情報等を含む。

記憶装置2は、オブジェクトの付加情報（例えば、意味情報。）と視覚的特徴情報（例えば、視覚的特徴量。）とを記憶する。意味情報とは、例えば、オブジェクトの名前や、歴史、利用料金、営業時間、住所、電話番号、URL等のオブジェクトに関する情報である。ここで、映像中のオブジェクトに意味情報を対応付けることを、オブジェクトにアノテーション（注釈）を付与するという。より具体的には、映像中のオブジェクトの近傍にアノテーション画像を付加することである。

視覚的特徴量とは、例えば、形状、大きさ、模様、エッジ、色等のオブジェクトを視覚的に認識するための特徴を数値化した量であって、例えば、山田昭雄、「ビジュアルツールによるコンテンツ記述」、映像情報メディア学会誌、社団法人映像情報メディア学会、2002年11月1日、第56巻、第11号、p. 1

711-1714（以下、文献Aと記す。）に記載されているように、MPEG-7/Visualで定義されているビジュアル記述子の色の特徴、模様の特徴、形状の特徴などのいずれかの組み合わせを用いて定義される特徴量ベクトルである。ここで、色の特徴はドミナントカラーとカラーレイアウト、模様の特徴はエッジヒストグラム、形状の特徴はリージョンベースド形状記述子等によって表される。

ドミナントカラーは、最頻色を頻度とともにRGB表記したものである。カラーレイアウトは、色の空間的な配置を周波数軸上で表現したものである。エッジヒストグラムは、例えば、縦や横、斜め等エッジがそれぞれどのくらい存在するかをヒストグラム化して記述したものである。リージョンベースド形状記述子は、2値の形状データを2次元ART（Angular Radial Transform）変換した35係数によって構成される記述子である。

オブジェクト $O_i$ の視覚的特徴量 $FO_i$ は、 $FO_i = (\text{ドミナントカラー、カラーレイアウト、エッジヒストグラム、リージョンベースド形状記述子、} \dots)$ と、ベクトル表記されて記述される。

記憶装置2は、一のオブジェクトに対して、撮影した日付と時刻とに応じて複数の視覚的特徴量を記憶してもよい。例えば一のオブジェクトに対して昼間の時刻の視覚的特徴量は明るい色を示したり、夜間の時刻の視覚的特徴情報は暗い色を示したりする。また、例えば山等の日付（例えば、季節等。）によって色が変化するオブジェクトに対して、冬は白色、春と夏とは緑色、秋は赤色や黄色等を示す視覚的特徴量とする等である。

記憶装置2は、例えば建築物や橋、塔等のような人工建築物や、例えば山や湖のような自然の地物等のように、経度と緯度等の情報である位置情報によって特定できるオブジェクトである場合、オブジェクトの位置情報を記憶する。以下、位置情報によって特定できるオブジェクトを、地理オブジェクトと記す。また、電車等のように時間の経過とともに移動するオブジェクトは、位置情報のみによって特定することはできないが、ダイヤグラム等により各時刻における位置が特定されれば、時刻と位置情報とを組み合わせた組み合わせ集合を用いてオブジェクトを特定してもよい。

位置情報に用いる位置の情報は、例えば地理オブジェクトの中心点または重心点等の緯度と経度とを用いてもよく、地理オブジェクトの形状を多角形（ポリゴン）の面を用いて構成した立体形状の各頂点の位置を、緯度と経度と標高とを用いて表したデータ群であってもよい。しかし、そのような立体形状の各頂点の位置のデータ群を位置情報に用いると、データ量が多くなってしまう。そこで、そのような立体形状の頂点のうち、1つまたは複数の頂点の位置情報を抽出して用いてもよい。また、地理オブジェクトの形状を円錐や、円柱、または立方体等の形状で近似し、近似した形状の中心点や、重心点、各頂点等の位置情報を用いてもよいし、1つの頂点の位置情報を抽出して用いてもよい。例えば、東京タワーを四角錐で近似し、各頂点の5つの位置情報を地理オブジェクトの位置情報に用いてもよいし、オブジェクトの最も標高の高い位置の頂点である最高点の位置情報を、地理オブジェクトの位置情報に用いてもよい。最高点は、撮影する場合に、他の頂点に比べて他の物体の存在によって隠れにくいと考えられるため、地理オブジェクトの最高点を地理オブジェクトの位置情報に用いることが好ましい。

意味情報と、最高点の緯度と経度と標高とを用いた位置情報と、視覚的特徴量との一例を図2に示す。視覚的特徴量は実際には数値化されているが、簡単のため形状と文字とで表した。

記憶装置2は、情報を、CD-ROMや、DVD-R、ハードディスク装置またはメモリ等の記録媒体に、複数に分割して、または一括して記憶する。記憶したデータは、読み込み装置が単数、または複数の媒体に記録された情報を読み込む。また、記憶装置2は、情報をインターネット等のコンピュータネットワークを介して取得してもよい。

オブジェクト認識装置3は、計算機で実現され、入力装置1が撮影した映像の一部または映像全体の視覚的特徴量と、記憶装置2が記憶している地理オブジェクトの視覚的特徴量との類似度を算出する。

図3に、本発明の第1の実施の形態のオブジェクト認識装置3の一構成例を示し、オブジェクト認識装置3の構成について説明する。なお、図3や図4等を示す各手段は、プログラムで実現される。

オブジェクト認識装置 3 は、撮影情報と地理オブジェクトの位置情報とに基づいて、撮影されている可能性のある地理オブジェクトである候補オブジェクトを抽出する撮影空間推定手段、マッチング手段および候補オブジェクト検索手段

(いずれも、例えば、候補オブジェクト検索手段 30 で実現される。) と、候補オブジェクト検索手段 30 が抽出した候補オブジェクトの視覚的特徴量と、入力装置 1 が撮影した映像の一部分の領域または映像全体の視覚的特徴量との類似度を算出する類似度算出手段 (例えば、視覚的特徴量比較手段 31 で実現される。) と、算出した類似度と、あらかじめ決められたしきい値とを比較して、オブジェクトが撮影されているか否かを判定する判定手段 32 とを含む。

図 4 に、視覚的特徴量比較手段 31 の一構成例を示し、構成について説明する。視覚的特徴量比較手段 31 は、入力装置 1 が撮影した映像の一部分または全体の映像である部分映像を抽出する部分映像抽出手段 310 と、抽出された部分映像の視覚的特徴量を設定する視覚的特徴量設定手段 311 と、記憶装置 2 が記憶している候補オブジェクトの視覚的特徴量と、部分映像の視覚的特徴量との類似度を算出する特徴量比較手段 312 とを含む。視覚的特徴情報設定手段は、視覚的特徴量設定手段 311 によって実現される。

次に、図 5 のフローチャートを参照して動作について説明する。まず、映像入力機器が映像を撮影し、入力装置 1 に映像を表す映像データを入力する。入力装置 1 は、入力された映像データを映像記録メディアに記録する (ステップ A1)。そして、入力装置 1 は、撮影位置や撮影方向、移動方向および移動速度の情報と、画角情報とを入力し、入力された各情報を組み合わせて撮影情報を生成し (ステップ A2)、生成した撮影情報を、映像が記録される映像記録メディアに記録する。ここで、映像の各コマの画像データ圧縮方式が J P E G である場合は画像データのヘッダに撮影情報を記録したり、または映像と撮影情報とをそれぞれ時刻に対応させたりして、映像と撮影情報とを 1 対 1 に対応させる。

一般的なビデオカメラ装置等の撮影装置は、一秒間に 30 コマ記録するが、GPS 機能等を用いて取得する経度と緯度との位置情報は、一秒間に 1 回程度しか取得できない。そこで、最後に取得した経度と緯度との位置情報を、撮影情報を



構成する移動方向および移動速度の情報を用いて補正して、補正した位置情報と映像とを1対1に対応させることが好ましい。

候補オブジェクト検索手段30は、入力装置1が生成した撮影情報に基づいて、映像入力装置が撮影可能な範囲である撮影空間を算出する（ステップA3）。

以下、撮影位置をC、撮影方向を $\theta$ とし、地理オブジェクトの代表点をO、地理オブジェクトの代表点Oの緯度を $O_{i1}$ 、経度を $O_{i2}$ 、標高を $O_{i3}$ として説明する。図6および図7に示すように、撮影空間は、撮影位置から撮影方向に対して、画角および地平面に囲まれた空間である。

図8に示すように、撮影位置をC、撮像素子の上右端、上左端、下右端および下左端をそれぞれ、UR、UL、DRおよびDLとし、CとURとULとを含む平面を $S_u$ 、CとURとDRとを含む平面を $S_r$ 、CとDRとDLとを含む平面を $S_d$ 、CとULとDLとを含む平面を $S_l$ 、地平面を $S_e$ とすると、それぞれの平面は、(1)式のように表すことができる。ここで、撮影方向 $\theta$ は地表面と平行な北の方向を0とし、(1)式は、平面 $S_u$ 、 $S_d$ 、 $S_r$ 、 $S_l$ 、 $S_e$ を、地表面の所定の位置を原点とする直交座標系で表わしたものある。

$$\begin{cases} S_u : f_{su}(x, y, z) = 0 \\ S_d : f_{sd}(x, y, z) = 0 \\ S_r : f_{sr}(x, y, z) = 0 \\ S_l : f_{sl}(x, y, z) = 0 \\ S_e : f_{se}(x, y, z) = 0 \\ \dots (1) \end{cases}$$

簡略化のため、画角が180度以内の場合について、説明する。この場合、撮影空間は、 $S_u$ と $S_d$ と $S_r$ と $S_l$ とに挟まれ、地平面 $S_e$ より上方であって、撮影方向 $\theta$ から90度以内の空間に限定される。そこで、撮影空間は、(2)式の条件を満たす空間として表すことができる。ここで、 $C_x$ と $C_y$ と $C_z$ とは、撮影位置Cの緯度 $C_1$ と経度 $C_2$ と標高 $C_3$ とを、それぞれ直交座標系に変換して求めた撮影位置Cの位置情報である。また、 $\theta_x$ は $\theta$ と緯線とが成す角であり、

$\theta_y$  は  $\theta$  と経線とが成す角であり、 $\theta_z$  は  $\theta$  と地平面に対する垂直線とが成す角である。

$$\left\{ \begin{array}{l} f_{su}(x, y, z) f_{sd}(x, y, z) \leq 0 \\ f_{sr}(x, y, z) f_{sl}(x, y, z) \leq 0 \\ f_{se}(x, y, z) \geq 0 \\ \begin{pmatrix} \theta_x \\ \theta_y \\ \theta_z \end{pmatrix} \cdot \begin{pmatrix} x - Cx \\ y - Cy \\ z - Cz \end{pmatrix} > 0 \end{array} \right.$$

. . . (2)

次に、オブジェクト認識装置 3 の候補オブジェクト検索手段 30 が、記憶装置 2 が記憶している地理オブジェクトの位置情報を読み出す（ステップ A 4）。候補オブジェクト検索手段 30 は、読み出した各地理オブジェクトの各位置情報を用いて、映像入力機器に撮影された可能性のある地理オブジェクトを検索する。映像入力機器に撮影された可能性のある地理オブジェクトとは、(2) 式を満たす位置情報を有する地理オブジェクトである。そこで、候補オブジェクト検索手段 30 は、(2) 式を満たす位置情報を有する地理オブジェクトである候補オブジェクトを抽出する。すなわち、撮影空間と、記憶装置 2 に記憶されている地理オブジェクトの位置情報が示す位置とのマッチングを行う。

(2) 式を満たす地理オブジェクトであっても、撮影位置から遠くに存在する地理オブジェクトは撮影されにくいため、撮影位置の情報と地理オブジェクトの位置情報とに基づいて算出した撮影位置と地理オブジェクトとの距離が、あらかじめ決められたしきい値  $Th1$  を超えている地理オブジェクトは抽出しなくてもよい。しかし、例えば高い山等の遠くからでも見ることができる地理オブジェクトは、撮影位置との距離が遠くても撮影される可能性がある。そこで、図 9 に示すように、撮影位置から地理オブジェクトの頂点を見上げる角度を求め、この角度があるしきい値  $Th2$  を超えている地理オブジェクトは抽出してもよい。

また、撮影方向  $\theta$  と、撮影位置から地理オブジェクトに向かう方向との差が大きい地理オブジェクトも撮影されにくい。従って、撮影位置 C から地理オブジェ

クトへの水平距離  $Dist$  がしきい値  $Th1$  以下であることと、撮影位置から地理オブジェクトの頂点を見上げる角度である仰角がしきい値  $Th2$  以下であることと、撮影方向  $\theta$  と撮影位置から地理オブジェクトに向かう方向との差である水平角度誤差がしきい値  $Th3$  以下であることとを、候補オブジェクトを抽出する条件に加えると候補オブジェクトの数を絞り込むことができる。そこで、 $Th1$ 、 $Th2$  および  $Th3$  を用いて、候補オブジェクトが満たす条件を (3) 式に示す。ここで、 $Oix$  と  $Oiy$  と  $Oiz$  とは、地理オブジェクトの代表点  $O$  の緯度  $Oi1$  と経度  $Oi2$  と標高を  $Oi3$  とをそれぞれ直交座標系に変換して求めた代表点  $O$  の位置情報である。この実施の形態では、(3) 式を用いて候補オブジェクトをさらに絞り込む (ステップ A5)。

$$\left\{ \begin{array}{l} Dist = \sqrt{(Cx - Oix)^2 + (Cy - Oiy)^2} \leq Th1 \\ \left| \theta - \tan^{-1} \left( \frac{Cx - Oix}{Cy - Oiy} \right) \right| \leq Th2 \\ \left| \tan^{-1} \left( \frac{Oiz - Cz}{Dist} \right) \right| \geq Th3 \end{array} \right.$$

... (3)

以下、(2) 式および (3) 式に示す条件を満たす  $M$  個の候補オブジェクトを、候補オブジェクト  $Oci$  ( $i = 1 \sim M$ ) として説明する。候補オブジェクト検索手段 30 は、候補オブジェクト  $Oci$  を抽出すると、抽出した候補オブジェクト  $Oci$  を特定する情報を視覚的特徴量比較手段 31 に出力する。

視覚的特徴量比較手段 31 は、記憶装置 2 があらかじめ記憶している候補オブジェクト  $Oci$  の視覚的特徴量  $FOi$  を読み込む (ステップ A6)。記憶装置 2 は、候補オブジェクト  $Oci$  の視覚的特徴量  $FOi$  を、 $FOi = (\text{ドミナントカラー, カラーレイアウト, エッジヒストグラム, リージョンベースド形状記述子, ...})$  のように、ベクトル表記して記憶している。

また、オブジェクトと類似する視覚的特徴を有するテンプレート映像を視覚的特徴量に用いてもよい。テンプレート映像とは、例えば、図 10 に示すような、

あらかじめ用意した一般的な形状のオブジェクトの映像である。そして、記憶装置 2 は、このようなテンプレート映像を視覚的特徴量として記憶していてもよい。

部分映像抽出手段 310 は、映像入力機器が撮影し、入力装置 1 が入力した映像データによる映像の一部分または全体を抽出した映像を表す部分映像データ

(以下、部分映像という。) を抽出する (ステップ A7)。部分映像の抽出方法の一例として、図 11 に示すように、長方形のブロックを映像の左上から右下まで、ラスタスキャンさせてブロック内の映像を抽出する方法がある。候補オブジェクトの視覚的特徴量に基づいて、または利用者の指示に基づいて、ブロックの大きさを変化させると、様々な大きさの部分映像を抽出することができる。また、ブロックの形状は長方形に限定されず、正方形や円形、楕円形等であってもよい。

また、部分映像抽出手段 310 は、候補オブジェクトの位置情報と、撮影情報を構成する撮影方向の情報を読み出して、各候補オブジェクト  $Oci$  の位置が撮影方向に対して右側であるか、左側であるか、中央付近であるかを推定してもよい。そして、部分映像の抽出のためにラスタスキャンする範囲を映像全体ではなく、映像の右半分の範囲や、左半分の範囲、右  $1/4$  の範囲、左  $1/4$  の範囲または中央付近等に限定することにより、部分映像の抽出に伴う部分映像抽出手段 310 の処理量を削減することができる。

図 12 に、撮影空間の右側に候補オブジェクトが位置する場合の説明図を示す。このとき、映像内で候補オブジェクトが存在する可能性のある範囲は、映像の右側に限定される。そこで、図 13 に示すように、部分映像を抽出する範囲を、映像の右側の範囲に限定してもよい。同様に、映像内で候補オブジェクトが存在する可能性のある範囲を上側の範囲あるいは下側の範囲等に特定して、部分映像を抽出する範囲を映像の上側の範囲あるいは下側の範囲に限定してもよいし、候補オブジェクトと撮影位置との距離や、候補オブジェクトの大きさ等によって抽出する部分映像の範囲を限定してもよい。

また、セグメンテーションを用いた部分映像を抽出方法を利用してもよい。セグメンテーションとは、画素および領域の色、エッジ、境界の複雑度等を分析して映像をいくつかの領域に分割することである。

セグメンテーションは、例えば、On image segmentation for object image retrieval、Hirata, K. ; Kasutani, E. ; Hara, Y.、Pattern Recognition、2002、Proceedings. 16th International Conference, on Volume 3、P. 1031-1034に記載の方法を利用して行われる。

セグメンテーションは、まず、隣接する同色の画素の集合による領域を生成する。そして、複雑な境界によって隣接し、互いに類似する色の領域を結合する。

次に、距離が近く、色と、模様とが類似する領域を結合する。そして、境界が複雑であって距離が近く、色と、模様とが類似する領域を結合する。最後に、広い領域の周辺に点在する狭い領域を、広い領域に結合する。

図14に領域に分割する映像の一例を示し、映像を領域に分割して抽出する方法について具体的に説明する。図14において、抽出する領域は、映像の右側に位置するBタワーであるものとする。

まず、セグメンテーションの方法を利用して、映像を、色およびエッジにより領域に分割する。領域に分割されたBタワーを図15の左の図に示す。分割された領域を $R_i$  ( $i=1\sim p$ ) とする。Bタワーは、 $R_4\sim R_7$ の領域に分割されたことが分かる。次に、セグメンテーションの方法を利用して、分割された領域を結合する。結合されたBタワーの領域の図を図15の右の図に示す。 $R_i$  と  $R_j$  とを結合した領域を $R_{i-j}$  とすると、結合されたBタワーの領域は $R_4-5-6-7$  となり、Bタワーの領域が結合されたことが分かる。

このように、領域 $R_i$  と領域 $R_j$  とを結合させた $R_{i-j}$  の生成を繰り返す。部分映像は、複数の領域同士の組み合わせの結合を行い、結合されたそれぞれの領域を抽出したものである。

また、領域の分割および領域の結合を効果的に行うため、候補オブジェクト毎に、領域を分割する方法と領域を結合する方法とを変更するとよい。以下、候補オブジェクトの視覚的特徴量を用いた領域の分割方法と結合方法との変更について、図16に映像の一例の図を示して説明する。図2のBタワーの欄に示すように、Bタワーの視覚的特徴量のエッジヒストグラムは、斜めの要素が大きいこと

が記述され、ドミナントカラーは「赤」および「白」の要素が強いことが記述されている。

そのため、領域の分割を行う際に、斜めエッジの境界線と、赤色の領域と、白色の領域とに対して感度を高めるように部分映像抽出手段 310 のパラメータを変更して分割を行うと、図 17 に示すように領域を分割することができる。図 17 において、実線部分が分割された領域の境界線を表している。このように、領域の分割と結合とを行うことにより、領域の数を減少させ、抽出する部分映像の数を減少させることができる。また、部分映像抽出手段 310 は、候補オブジェクトの視覚的特徴量に応じて、例えば、輝度情報や、色情報、形状の情報、模様の情報、大きさの情報等の情報のうちの 1 つの条件、または複数を組み合わせた条件に合致する映像の領域の部分映像を抽出してもよい。

部分映像抽出手段 310 は、抽出した部分映像を、視覚的特徴量設定手段 311 に出力する。視覚的特徴量設定手段 311 は、部分映像抽出手段 310 が抽出した部分映像の視覚的特徴量を、公知の種々の方法のいずれかを用いて例えば前述した文献 A に記載の方法で算出して設定する（ステップ A8）。そして、視覚的特徴量設定手段 311 は、設定した視覚的特徴量を特徴量比較手段 312 に出力する。

特徴量比較手段 312 は、候補オブジェクトの視覚的特徴量  $F_{O_i}$  と部分映像の視覚的特徴量  $F_r$  との類似度  $MS_i$  を算出する。図 18 に、部分映像と候補オブジェクトとを比較する例を示す。ここで、候補オブジェクトの視覚的特徴量  $F_{O_i}$  は、 $F_{O_i} = (\text{ドミナントカラー, カラーレイアウト, エッジヒストグラム, リージョンベースド形状記述子, } \dots)$  で表されるものとし、部分映像の視覚的特徴量  $F_r$  は、 $F_r = (\text{ドミナントカラー, カラーレイアウト, エッジヒストグラム, リージョンベースド形状記述子, } \dots)$  で表されるものとする。そして、類似度ベクトル  $S_i$  は、 $S_i = (\text{ドミナントカラーの類似度, カラーレイアウトの類似度, エッジヒストグラムの類似度, リージョンベースド形状記述子の類似度, } \dots)$  で表されるものとする。

ここで、ドミナントカラーの類似度は、候補オブジェクトの代表色と部分映像の代表色との相関を考慮して算出される。カラーレイアウトの類似度は、色成分

毎と、周波数成分毎との視覚特性を反映する重み付けを、輝度成分係数と色差成分係数とに導入して算出したL2ノルム和である。エッジヒストグラムの類似度は、例えば5個のエッジ要素による、候補オブジェクトのヒストグラムと部分映像のヒストグラムとにおける各エッジ要素の差分絶対値和である。リージョンベースド形状記述子の類似度は、記述子を構成する要素毎に算出した、部分映像と候補オブジェクトとの差分絶対値和である。

すると、類似度 $MS_i$ は(4)式のように表し、類似度を算出する(ステップA9)。

$$MS_i = |S_i|$$

. . . (4)

ただし、類似度ベクトル $S_i$ の成分である各類似度は、0以上で1以下になるように正規化しておくものとする。

また、視覚的特徴量比較手段31が、テンプレート映像を視覚的特徴量に用いた場合の類似度の計算方法について説明する。テンプレート映像を $T_{oi}(x, y)$ 、映像を $F(x, y)$ テンプレート映像の面積を $SF$ とすると、類似度 $MS_i(X, Y)$ は(5)式のように表し、算出することができる。

$$MS_i(X, Y) = \sum_{x, y} \frac{|F(X+x, Y+y) - T_{oi}(x, y)|}{SF}$$

. . . (5)

テンプレート映像の範囲を、例えば $0 \leq x \leq 29$ と $0 \leq y \leq 49$ とすると、(5)式は、映像上のある一点 $(X, Y)$ からテンプレート映像の範囲の領域の映像と、テンプレート映像との類似度を表している。

以上の類似度の算出を、映像全体または部分映像抽出手段310が抽出したすべての部分映像について行い、算出した類似度を判定手段32に出力する。ここ

で、映像全体または部分映像と、候補オブジェクトとの関係は、算出した類似度の数値が小さいほど視覚的な類似性が強いことを表す。そこで、以下、算出した類似度の数値が最も少ないものを、最大類似度と記す。

判定手段 32 は、候補オブジェクトに対する最大の類似度  $MS_{max i}$  を抽出して最大類似度  $MS_{max i}$  があらかじめ決められたしきい値  $Th_i$  より小さいか否かの判定を行う（ステップ A11）。

最大類似度  $MS_{max i}$  がしきい値  $Th_i$  より小さい場合、判定手段 32 は、候補オブジェクトが映像中に存在すると判断して、候補オブジェクトである地理オブジェクトが撮影されている部分映像と、候補オブジェクトの意味情報と記憶装置 2 から読み出し、対応付けて出力する（ステップ A12）。つまり、映像にアノテーションを付与する。出力先は、例えば液晶表示装置等の表示装置、ハードディスク装置やメモリ等の記憶装置等である。

なお、図 19 に示すような、例えば、お寺等の複数の建造物からなる地理オブジェクトでは、お寺全体の視覚的特徴量が見る位置により大きく異なることがある。そこで、地理オブジェクトである個々の建造物（例えば、本堂と五重塔等。）の地理オブジェクトに対応する意味情報を階層的に（例えばツリー状に）記憶装置 2 に記憶させておくことが好ましい。例えば、複数の意味情報に共通する概念（共通概念）が存在する場合に、それらの意味情報は、ツリーの 1 つのノードから延びる複数の枝に相当する意味情報として記憶装置 2 に記憶される。そして、ツリーのノードに相当する共通意味情報（各意味情報の共通概念を示す意味情報）も、記憶装置 2 に記憶させておくことが好ましい。さらに、複数の共通意味情報に共通する概念が存在する場合に、それらの共通意味情報は、ツリーの 1 つのノードから延びる複数の枝に相当する上位層の共通意味情報として記憶装置 2 に記憶される。そして、共通意味情報のノードに相当する上位層の共通意味情報も、記憶装置 2 に記憶させておくことが好ましい。すなわち、図 20 に示すように、地理オブジェクトの集合（例えば、本堂と五重塔。）の階層的な上位概念の意味情報（例えば、D 寺。）をあらかじめ作成しておく。そして、判定手段 32 は、図 19 中に円で示したように、抽出された地理オブジェクトについて、共通意味情報の下位概念に相当する意味情報（例えば、本堂または五重塔。）に



対応する地理オブジェクトを認識すると、地理オブジェクトの集合の上位概念として「D寺」というアノテーションを付与する。

同様に、例えば、図21に示すCビルディング等の地理オブジェクトの輪郭は、撮影方向によって大きく異なる場合がある。図22に示すように、そのような地理オブジェクトの構成を分割し、分割された各々の地理オブジェクトの視覚的特徴量をそれぞれ記憶装置2に記憶させてもよい。その場合、図20に示すように、分割された各々の地理オブジェクトの意味情報の上位概念に相当する共通意味情報（例えば、「Cビルディング」）を記憶装置2に記憶させ、さらに、共通意味情報の上位概念に相当する上位層の共通意味情報（例えば、「新宿副都心ビルディング群」）があれば、それも記憶装置2に記憶させておく。部分映像抽出手段310が、分割された地理オブジェクトの各部分を別個の地理オブジェクトとして抽出した場合に、判定手段32は、抽出された地理オブジェクトについて、意味情報（例えば、「Cビルディング右」）の上位概念である共通意味情報が記憶手段2に記憶されているか否か判定する。そして、共通意味情報が記憶手段2にあれば、地理オブジェクトの集合の上位概念として「Cビルディング」というアノテーションを付与する。

例えば木や人等の障害物によって地理オブジェクトの一部が隠れてしまう場合や、複雑な形状の地理オブジェクトの一部分によって他の大部分が隠れてしまう場合のように、地理オブジェクトの全体を撮影されていない映像が入力装置1に入力された場合がある。そのような場合、地理オブジェクトの一部を分割してそれぞれの地理オブジェクトの視覚的特徴量を記憶装置2に記憶させておけば、部分映像抽出手段310は、撮影されている一部の部分によって地理オブジェクトを認識して部分映像を抽出することができる。すると、地理オブジェクトを分割しないで部分映像を抽出する場合と比べて、建物の向きの変動に強いアノテーションの付与を行うことができる。

また、地理オブジェクトの全体のみならず、これに加えて、またはこれに代えて、一部が隠れた地理オブジェクトの視覚的特徴量や、一方向のみならず、これに加えて、またはこれに代えて、様々な方向・位置から撮影した地理オブジェクトの視覚的特徴量を複数あらかじめ記憶装置2に記憶させておいてもよい。これ

により入力映像の撮影位置、または、撮影方向に最も近い方向から撮影された際の特徴量を用いることにより見る向きにより視覚的特徴が大きく異なる地理オブジェクトにも対応することができる。

以上の、ステップA 6からステップA 12までの各ステップを、すべての候補オブジェクトについて繰り返し行う（ステップA 13）。本発明による映像アノテーション付与装置は、入力装置1に入力された映像の一コマ一コマに対して、ステップA 1からステップA 13までの処理を行う。しかし、例えば、映像入力機器が撮影する映像が一秒間に30コマであって、入力装置1に入力される映像が一秒間に3コマである場合には、映像入力機器が撮影した10コマの映像のうち、1コマの映像に対してのみアノテーションの付与が行われる。そこで、そのような場合には、アノテーションの出力を1/3秒間保持してもよい。そのような構成によれば、映像入力機器が撮影した映像に対して途切れることなくアノテーションの付与を行うことができる。

この実施の形態では、部分映像の視覚的特徴量と、記憶装置2が記憶しているオブジェクトの視覚的特徴量とを比較して、オブジェクトが映像に映っているか否かの判定を行う。そして、オブジェクトが映像に映っていると判定した場合に、オブジェクトとアノテーションとを対応付けるため、実際には映像に映っていないオブジェクトのアノテーションと映像とを対応付けることを防ぐことができる。また、部分映像との類似度を算出する地理オブジェクトを、記憶装置2が記憶している地理オブジェクトのうち、撮影空間に存在するか否か等の条件に合致する地理オブジェクトに絞り込んでいる。そのため、視覚的特徴量比較手段31のデータ処理の負担を軽減させることができる。

#### （発明の第2の実施の形態）

図23に、本発明の第2の実施の形態を説明するブロック図を示し、この実施の形態の構成について説明する。本発明の第2の実施の形態は、映像および映像を撮影した状況の情報である撮影情報が入力される入力装置1と、映像の被写体であるオブジェクトの情報を記憶する記憶装置2と、撮影情報とオブジェクトの情報とを比較して、オブジェクトが映像に撮影されているか否かを判定するオブジェクト認識装置4とを含む。

入力装置 1 と記憶装置 2 とは、第 1 の実施の形態と同様なため、説明を省略する。オブジェクト認識装置 4 は、撮影情報を構成する位置情報と地理オブジェクトの位置情報とに基づいて、地理オブジェクトが撮影されている確率である存在確率を算出する撮影空間推定手段、マッチング手段および存在確率算出手段（例えば、存在確率算出手段 40。）と、記憶装置 2 が記憶している地理オブジェクトの視覚的特徴量と、入力装置 1 に入力された映像の一部分の領域または全体の視覚的特徴量との類似度を算出する類似度算出手段（例えば、視覚的特徴量比較手段 41。）と、存在確率と類似度とを総合的に評価して、地理オブジェクトが撮影されているか否かを判断する判定手段 42 とを含む。

図 24 に、視覚的特徴量比較手段 41 の一構成例を示し、構成について説明する。視覚的特徴量比較手段 41 は、入力装置 1 に入力された映像から、部分映像を抽出する部分映像抽出手段 410 と、抽出した部分映像の視覚的特徴量を設定する視覚的特徴量抽出 411 と、記憶装置 2 が記憶している地理オブジェクトの視覚的特徴量と、部分映像の視覚的特徴量との類似度を算出する特徴量比較手段 412 とを含む。

次に、この実施の形態の動作について説明する。この実施の形態の動作と第 1 の実施の形態の動作との違いは、図 5 に示すフローチャートにおけるステップ A5 とステップ A11 とである。そのため、その他の動作については説明を省略する。

図 5 に示すフローチャートのステップ A5 において、第 1 の実施の形態では候補オブジェクト検索手段 30 が、候補オブジェクトの抽出を行う。しかし、この実施の形態では、候補オブジェクトの抽出を行わない。従って、特徴量比較手段 412 は、記憶装置 2 が記憶している全ての地理オブジェクトの視覚的特徴量と部分映像の視覚的特徴量との比較を行う。

図 5 に示すフローチャートのステップ A11 において、第 1 の実施の形態では最大類似度  $MS_{max i}$  のしきい値判定を行う。しかし、この実施の形態では、存在確率算出手段 40 が、記憶装置 2 が記憶している各地理オブジェクトが撮影されている確率である存在確率  $P_i$  を算出する。そして、判定手段 42 が、存在確率  $P_i$  と特徴量比較手段 412 が算出した最大類似度  $MS_{max i}$  の逆数とを

乗じて統合スコアを算出する。算出した統合スコアの数値が、あらかじめ決められたしきい値より大きい場合、判定手段 42 は、地理オブジェクトが映像中に存在すると判断して記憶装置 2 から地理オブジェクトの意味情報を読み出す。そして、地理オブジェクトが撮影されている部分映像と、記憶装置 2 から読み出した地理オブジェクトの意味情報とを対応付けて出力する。

存在確率  $P_i$  を求めるため、(3) 式の水平距離、水平角度誤差および仰角に対して、地理オブジェクトが撮影されている確率の分布である確率密度分布をあらかじめ作成し、これらの確率密度分布に基づいて存在確率を算出する。ここで、確率密度分布は、横軸を画素数、縦軸を確率としたガウス分布である。地理オブジェクトの映像上の位置の誤差は、GPS の計測精度と、撮影方向の角度の精度と、撮影位置から地理オブジェクトまでの距離とに基づいて誤差が推定できるので、その誤差をガウス分布の  $\sigma$  値とする等して確率密度関数を決定する。

この実施の形態では、地理オブジェクトが撮影されている確率と、部分映像の視覚的特徴量と、記憶装置 2 が記憶しているオブジェクトの視覚的特徴量とを比較した結果とに基づいて、オブジェクトが映像に映っているか否かの判定を行う。そのため、撮影空間に存在するオブジェクトであっても、映像に写っている確率が低いオブジェクトのアノテーションを、映像に対応付けることを防ぐことができる。

#### (発明の第 3 の実施の形態)

図 25 に、本発明の第 3 の実施の形態を説明するブロック図を示し、この実施の形態の構成について説明する。本発明の第 3 の実施の形態は、第 1 の実施の形態の構成に、レイアウト装置 51 と、表示装置 52 とを加えた構成である。そのため、第 1 の実施の形態と同様な装置等には図 1 と同様な符号を用いて、説明を省略する。表示位置決定手段は、レイアウト装置 51 によって実現される。

レイアウト装置 51 は、映像と、映像から抽出した部分映像に付与したアノテーションとを重畳して表示させる場合のアノテーションの表示位置を決定して、表示装置 52 に出力する。表示装置 52 は、映像にアノテーションを重畳し、レイアウト装置 51 が決定した表示位置にアノテーションを表示する。

図 26 は、表示装置 52 がアノテーションを映像に重畳させて表示した場合の表示例である。レイアウト装置 51 は、アノテーションの表示位置を、部分映像の真下、真上、右、または左のいずれかの位置であって、他の部分映像に重ならない位置に決定する。図 26 のように、他の部分映像が存在するために、所定の距離を超えて部分映像から離れた位置にアノテーションを表示位置させるときは、部分映像とアノテーションとを矢印でつないで表示させるとよい。

この実施の形態では、アノテーションを表示させる映像上の位置を決定する。そのため、オブジェクトとアノテーションとの対応を映像で表すことができる。

#### (発明の第 4 の実施の形態)

図 27 に、本発明の第 4 の実施の形態を説明するブロック図を示し、この実施の形態の構成について説明する。本発明の第 4 の実施の形態は、第 1 の実施の形態の構成に、アノテーション結果記憶装置 61 を加えた構成である。そのため、第 1 の実施の形態と同様な装置等には、図 1 と同様な符号を用いて、説明を省略する。

アノテーション結果記憶装置 61 は、部分映像と、部分映像に対応付けられた地理オブジェクトの意味情報との組を、対応を保持したまま記憶する。記憶媒体は、例えばメモリ、ハードディスク装置、CD-ROM、またはDVD-RAM 等である。また、アノテーション結果記憶装置 61 は、部分映像の映像上の位置の情報を記憶してもよく、入力されたキーワードに合致する映像を出力してもよい。例えば、全体の映像における部分映像の位置を、 $x$  座標と  $y$  座標との数値で特定できる場合であって、アノテーション結果記憶装置 61 が、B タワーが左側に写っている映像と、意味情報 (B タワー) と、全体の映像における部分映像の位置 ( $(x, y) = (120, 240)$ ) の情報の組を対応させて記憶した場合、「オブジェクト=B タワー」と、「領域= $x < 320$ 」とのキーワードが入力されると、以下のように映像が抽出される。ここで、全体の映像の  $x$  の最大値は 640 であったとする。

まず、意味情報が「B タワー」であって、全体の映像における部分映像の位置が「 $x < 320$ 」であるという条件に合致する意味情報とオブジェクトの領域との組を検索して特定する。すると、特定した組に対応付けられた映像である「B

タワーが左側に写っている映像」を抽出することができる。利用者はアノテーション結果記憶装置 61 が備えるキーボード等を介してキーワードを入力してもよいし、マイクロフォン等を介して音声入力を行ってもよい。また、利用者があらかじめアノテーション結果記憶装置 61 に自己の嗜好を記憶させ、アノテーション結果記憶装置 61 が記憶している利用者の嗜好に基づいてキーワードを生成してもよい。

この実施の形態は、アノテーションと、映像データに含まれるオブジェクトとを対応させて記憶するアノテーション結果記憶手段を備える。そのため、記憶したアノテーションと映像データに含まれるオブジェクトとの対応させに基づいて、映像データの検索を行うことができる。

#### (発明の第 5 の実施の形態)

図 28 に本発明の第 5 の実施の形態における視覚的特徴量比較手段を説明するブロック図を示し、この実施の形態の構成について説明する。

本発明の第 5 の実施の形態は、第 1 の実施の形態における視覚的特徴量比較手段の図 4 の構成に、結果保持手段 513 を加えた構成である。結果保持手段 513 は、特徴量比較手段 512 で算出した類似度を記録し類似度が高い部分映像のみを抽出する。

次に、この実施の形態の動作について説明する。この実施の形態の動作と第 1 の実施の形態の動作との違いは、図 29 に示すフローチャートにおけるステップ A101 からステップ A106 とである。そのため、その他の動作については説明を省略する。

図 29 に示すフローチャートのステップ A101 において、入力映像をセグメンテーションにより分割する。この分割された領域の組み合わせを結合して領域を得る。しかし、組み合わせ数は (2 の領域数乗) -1 であるため、領域数  $b$  の増加により単純組み合わせ数が指数関数的に増加してしまう。そこで、組み合わせの評価を階層的に行い絞り込む方法を導入する。例えば、図 30 に示すように、階層の 1 段目では、1 領域と対象ランドマークとの類似度を計算し、上位  $m$  個の領域のみを次段に引き渡し、 $m+1$  位以下の組み合わせについては今後評価対象としない。 $n$  段階目では、 $n-1$  段目で抽出された  $n-1$  領域の組み合わせに対し、新

たな1領域を加えたn個の領域を組み合わせ（ステップA102）、特徴量を抽出し（ステップA103）、対象ランドマークとの類似度を計算し（ステップA104）、上位m個の組み合わせを抽出する（ステップA107）。この処理をL(< b)段まで行い（ステップA106）、最大類似度となる組み合わせ領域を出力する。これにより、組み合わせ数を最大  $a(b(1+m)-m)$  と比例オーダに抑えることができる。このように視覚的に分割された領域を組み合わせていくことにより、対象ランドマークと異なる視覚的特徴をもつ隠蔽物を取り除いたランドマーク領域を抽出することが出来る。

このとき、組み合わせ段数に応じて、類似度ベクトル  $S_i$  計算の際に特徴量の重み付けを変更してもよい。例えば、組み合わせの段数が少ない場合は、組み合わせ領域は対象ランドマークの一部でしかないと考えられるため、類似度ベクトル  $S_i$  計算の際に形状に関する特徴量（リージョンベースド形状記述子）の類似度に小さな係数を重み付けし、形状に関する類似度の効果を低減させ、組み合わせの段数が多い場合は、類似度に大きな係数を重み付けし、形状の効果を増大させることで、効果的な絞込みを行える。

また、以上に述べた本発明による各実施の形態の構成は、車両案内システム（カーナビゲーションシステム）に搭載されて実現されてもよい。

以上に述べた本発明による各実施の形態は、撮影位置とオブジェクトとの距離の計測を行わずに、撮影されたオブジェクトと記憶装置が記憶しているオブジェクトとの対応付けを行う。撮影位置とオブジェクトとの距離に基づいて、撮影されたオブジェクトと記憶装置が記憶しているオブジェクトとの対応付けを行う方法がある。その場合に行われる撮影位置とオブジェクトとの距離の計測は、複数のカメラを用いたステレオ視の方法や、レーザ等を用いた光学的距離測定方法がある。しかし、これらの方法を用いた距離の測定には、複数のカメラあるいは光学式距離測定装置を必要とし、距離の測定に時間を要する。また、オブジェクト付近の木や看板等と、オブジェクトとの距離による識別は困難であるという問題がある。

### 請求の範囲

1. 映像データと、撮影空間を決定しうる情報である撮影情報とを入力する入力手段と、

オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶する記憶手段と、

入力された映像データによる映像に含まれるオブジェクトを認識するオブジェクト認識手段とを備え、

前記オブジェクト認識手段は、

撮影情報に基づいて撮影空間を推定する撮影空間推定手段と、

撮影空間と、前記記憶手段に記憶されているオブジェクトの位置情報が示す位置とのマッチングを行うマッチング手段と、

入力された映像データから、映像データによる映像の一部分の領域の映像データまたは映像全体の映像データである部分映像データを抽出する部分映像抽出手段と、

部分映像データの視覚的特徴情報を生成する視覚的特徴情報設定手段と、

部分映像データの視覚的特徴情報と、前記記憶手段が記憶しているオブジェクトの視覚的特徴情報とを比較して類似度を算出する類似度算出手段と、

前記マッチング手段のマッチングの結果と、算出した類似度とに基づいて、入力された映像データによる映像中にオブジェクトが存在しているか否かを判定する判定手段とを備えた映像オブジェクト認識装置。

2. 映像データと、撮影空間を決定しうる情報である撮影情報とを入力する入力手段と、

オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶する記憶手段と、

入力された映像データによる映像に含まれるオブジェクトと、付加情報とを対応付けるオブジェクト認識手段とを備え、

前記オブジェクト認識手段は、

撮影情報に基づいて撮影空間を推定する撮影空間推定手段と、



撮影空間と、前記記憶手段に記憶されているオブジェクトの位置情報が示す位置とのマッチングを行うマッチング手段と、

入力された映像データから、映像データによる映像の一部分の領域の映像データまたは映像全体の映像データである部分映像データを抽出する部分映像抽出手段と、

部分映像データの視覚的特徴情報を生成する視覚的特徴情報設定手段と、

部分映像データの視覚的特徴情報と、前記記憶手段が記憶しているオブジェクトの視覚的特徴情報とを比較して類似度を算出する類似度算出手段と、

前記マッチング手段のマッチングの結果と、算出した類似度とに基づいて、入力された映像データによる映像に含まれるオブジェクトを特定し、特定したオブジェクトと、前記記憶手段が記憶している付加情報とを対応付ける判定手段とを備えた映像アノテーション付与装置。

3. 前記オブジェクト認識手段は、

撮影空間と、記憶手段が記憶しているオブジェクトの位置情報とに基づいてオブジェクトが映像に含まれている確率である存在確率を算出する存在確率算出手段を含み、

前記判定手段は、算出した存在確率と類似度とに基づいて映像に含まれるオブジェクトを特定し、特定したオブジェクトと前記記憶手段が記憶している付加情報とを対応付ける

請求の範囲第2項に記載の映像アノテーション付与装置。

4. 前記部分映像抽出手段は、

前記記憶手段が記憶しているオブジェクトの位置情報に基づいて、映像におけるオブジェクトが位置する範囲を特定し、特定した範囲から部分映像データを抽出する

請求の範囲第3項に記載の映像アノテーション付与装置。

5. 前記オブジェクト認識手段は、

撮影空間と位置情報とに基づいて撮影空間に存在するオブジェクトである候補オブジェクトの抽出を行う候補オブジェクト検索手段を含み、

前記類似度算出手段は、部分映像データの視覚的特徴情報と、記憶手段が記憶している候補オブジェクトの視覚的特徴情報とを比較して類似度を算出する

請求の範囲第2項に記載の映像アノテーション付与装置。

6. 前記部分映像抽出手段は、

前記記憶手段が記憶している候補オブジェクトの位置情報に基づいて、映像におけるオブジェクトが位置する範囲を特定し、特定した範囲の映像から部分映像データを抽出する

請求の範囲第5項に記載の映像アノテーション付与装置。

7. 映像を表示する表示手段と、

前記表示手段に、映像に含まれるオブジェクトに対応付けられた付加情報を表示させる位置を指定して付加情報を映像に重畳して表示させる表示位置決定手段とを備えた

請求の範囲第2項に記載の映像アノテーション付与装置。

8. 付加情報と映像に含まれるオブジェクトとを対応させて記憶するアノテーション結果記憶手段を備えた

請求の範囲第2項に記載の映像アノテーション付与装置。

9. 前記部分映像抽出手段は、

抽出する部分映像データによる映像の領域の形状と大きさを任意に変更可能な機能を有する

請求の範囲第2項に記載の映像アノテーション付与装置。

10. 前記部分映像抽出手段は、

輝度情報と、色情報と、形状の情報と、模様の情報と、大きさの情報とのうちの1つの条件、または複数を組み合わせた条件に合致する映像の領域の部分映像データを抽出する

請求の範囲第2項に記載の映像アノテーション付与装置。

11. 前記部分映像抽出手段は、

複数の情報を組み合わせた条件に合致する映像の領域の部分映像データを抽出する場合、マッチング手段のマッチングの結果と、記憶手段が記憶しているオブ

ジェクトの視覚的特徴情報とに基づいて、各条件に対する重み付けを行って部分映像データを抽出する

請求の範囲第10項に記載の映像アノテーション付与装置。

12. 前記記憶手段が記憶するオブジェクトの視覚的特徴情報は、オブジェクトに類似する視覚的な特徴を有する映像であるテンプレート映像である

請求の範囲第2項に記載の映像アノテーション付与装置。

13. 前記記憶手段が記憶するオブジェクトの視覚的特徴情報は、色情報、形状の情報、模様の情報および大きさの情報のうち1つ、または複数の情報によって構成され、

前記視覚的特徴情報設定手段が生成する部分映像データの視覚的特徴情報は、色情報と、形状の情報と、模様の情報と、大きさの情報とのうち1つの情報、または複数の情報によって構成される

請求の範囲第2項に記載の映像アノテーション付与装置。

14. 前記記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの形状に類似した形状の円錐、円柱、または立方体等の立体幾何学の立体形状を用いて近似した立体形状の頂点のうちの1つの頂点、または中心点、もしくは重心点のいずれかの位置を特定する情報である

請求の範囲第2項に記載の映像アノテーション付与装置。

15. 前記記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの形状を多角形の面を用いて近似した三次元形状の頂点のうち少なくとも1つの頂点の位置を特定する情報である

請求の範囲第2項に記載の映像アノテーション付与装置。

16. 前記記憶手段が記憶するオブジェクトの位置情報は、オブジェクトの頂点のうち、最も標高が高い頂点の位置を特定する情報である

請求の範囲第2項に記載の映像アノテーション付与装置。

17. 前記記憶手段が記憶するオブジェクトの位置情報は、緯度と経度と標高とによってオブジェクトの位置を特定する情報である

請求の範囲第2項に記載の映像アノテーション付与装置。

18. 前記記憶手段は、複数のオブジェクトのそれぞれに対応した付加情報に共通する概念にもとづく共通付加情報、または複数の共通付加情報に共通する概念にもとづく共通付加情報を階層的に記憶し、

前記判定手段は、撮影されたオブジェクトの付加情報または共通付加情報に対応する共通付加情報が存在するか否かを判定し、存在する場合には、そのオブジェクトに共通付加情報を対応付ける

請求の範囲第2項に記載の映像アノテーション付与装置。

19. 前記撮影情報は、撮影した日付と時刻とを特定する情報である撮影日時情報を含み、

前記記憶手段は、撮影した日付と、時刻とに応じた視覚的特徴情報を記憶し、

前記類似度算出手段は、部分映像データの視覚的特徴情報と、撮影日時情報が特定する日付と、時刻とに応じた視覚的特徴情報とを比較して類似度を算出する

請求の範囲第2項に記載の映像アノテーション付与装置。

20. 前記部分映像抽出手段は、

前記入力された映像データから領域を分割し、分割された領域を前記部分映像データとする

請求の範囲第10項に記載の映像アノテーション付与装置。

21. 前記部分映像抽出手段は、

前記分割された領域の組み合わせにより前記部分映像データとする

請求の範囲第20項に記載の映像アノテーション付与装置。

22. 前記部分映像抽出手段は、

前記分割領域の組み合わせの評価を階層的に行うことで部分映像データを生成する

請求の範囲第21項に記載の映像アノテーション付与装置。

23. 前記部分映像抽出手段は、

前記分割領域の組み合わせ階層評価として、同一領域数の組み合わせから前記類似度の高い特定数のみ以後の組み合わせに用いる

請求の範囲第22項に記載の映像アノテーション付与装置。

24. 前記記憶手段が記憶するオブジェクトの視覚的特徴情報として、

一方向または、複数方向からの全体像または一部分のオブジェクトにおける視覚的情報を複数保持する

請求の範囲第2項に記載の映像アノテーション付与装置。

25. 車両に搭載され、GPSを用いて自己の位置を表示装置が表示する地図上に示す車両案内システムであって、

請求の範囲第2項に記載の映像アノテーション付与装置を備えた車両案内システム。

26. 映像データと、撮影空間を決定しうる情報である撮影情報とを入力し、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶し、

撮影情報に基づいて撮影空間を推定し、

撮影空間と、記憶されているオブジェクトの位置情報が示す位置とのマッチングを行い、

入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出し、

部分映像データの視覚的特徴情報を生成し、

部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出し、

マッチングの結果と、算出した類似度とに基づいてオブジェクトが撮影されているか否かを判定する映像オブジェクト認識方法。

27. 映像データと、撮影空間を決定しうる情報である撮影情報とを入力し、オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶し、

撮影情報に基づいて撮影空間を推定し、

撮影空間と、記憶されているオブジェクトの位置情報が示す位置とのマッチングを行い、

入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出し、

部分映像データの視覚的特徴情報を生成し、

部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出し、

マッチングの結果と、算出した類似度とに基づいて映像データに含まれるオブジェクトを特定し、特定したオブジェクトと記憶されている付加情報によって構成されるアノテーションとを対応付ける映像アノテーション付与方法。

28. 映像データに、記憶されているオブジェクトが被写体として含まれているか否かを判定する映像オブジェクト認識装置に搭載される映像オブジェクト認識プログラムであって、

コンピュータに、

オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報とを対応させて記憶装置に記憶させる処理と、

撮影空間を決定しうる情報である撮影情報に基づいて撮影空間を推定する処理と、

撮影空間と、前記記憶装置が記憶しているオブジェクトの位置情報が示す位置とのマッチングを行う処理と、

入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出する処理と、

部分映像データの視覚的特徴情報を生成する処理と、

部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出する処理と、

マッチングの結果と、算出した類似度とに基づいてオブジェクトが撮影されているか否かを判定する処理とを実行させる映像オブジェクト認識プログラム。

29. オブジェクトと、記憶されているオブジェクトの情報とを対応付ける映像アノテーション付与装置に搭載される映像アノテーション付与プログラムであって、

コンピュータに、

オブジェクトの位置の情報である位置情報と、オブジェクトの視覚的な特徴を数値化した情報である視覚的特徴情報と、オブジェクトに付加される情報である付加情報とを対応させて記憶装置に記憶させる処理と、

撮影空間を決定しうる情報である撮影情報に基づいて撮影空間を推定する処理と、

撮影空間と、前記記憶装置が記憶しているオブジェクトの位置情報が示す位置とのマッチングを行う処理と、

入力された映像データから、映像データによる映像の一部分の領域または映像全体の映像データである部分映像データを抽出する処理と、

部分映像データの視覚的特徴情報を生成する処理と、

部分映像データの視覚的特徴情報と、記憶されているオブジェクトの視覚的特徴情報とを比較して類似度を算出する処理と、

マッチングの結果と、算出した類似度とに基づいて映像データに含まれるオブジェクトを特定し、特定したオブジェクトと記憶されている付加情報によって構成されるアノテーションとを対応付ける処理とを実行させる映像アノテーション付与プログラム。

図 1

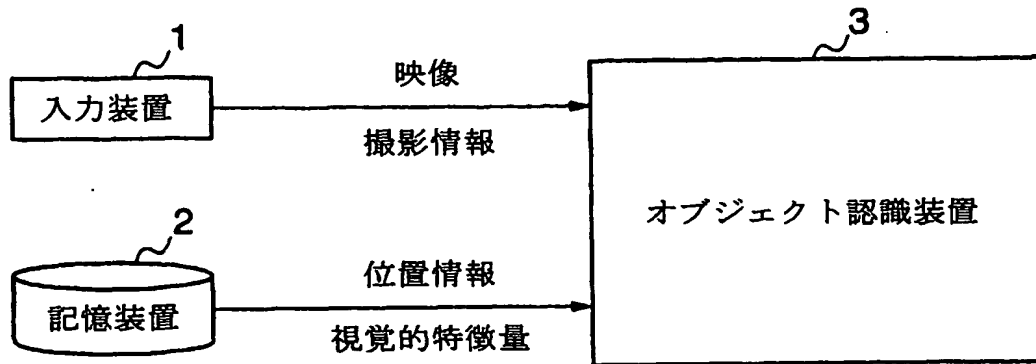


図 2

意味情報	位置情報	視覚的特徴量
A 山	(E138° 44' 02.03", N35° 21' 34.35", 3776m),	(△白青etc.)
B タワー	(E139° 44' 55.38", N35° 39' 18.59", 333m),	(△赤白etc.)
C ビル ディング	(E139° 45' 04.39", N35° 38' 46.16", 180m),	(□灰色etc.)



図 3

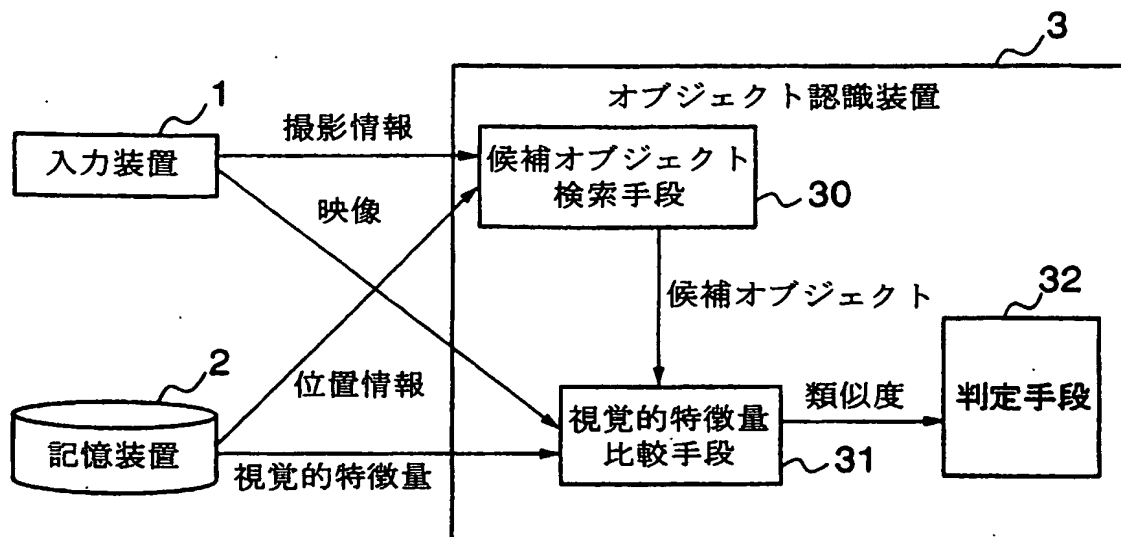


図 4

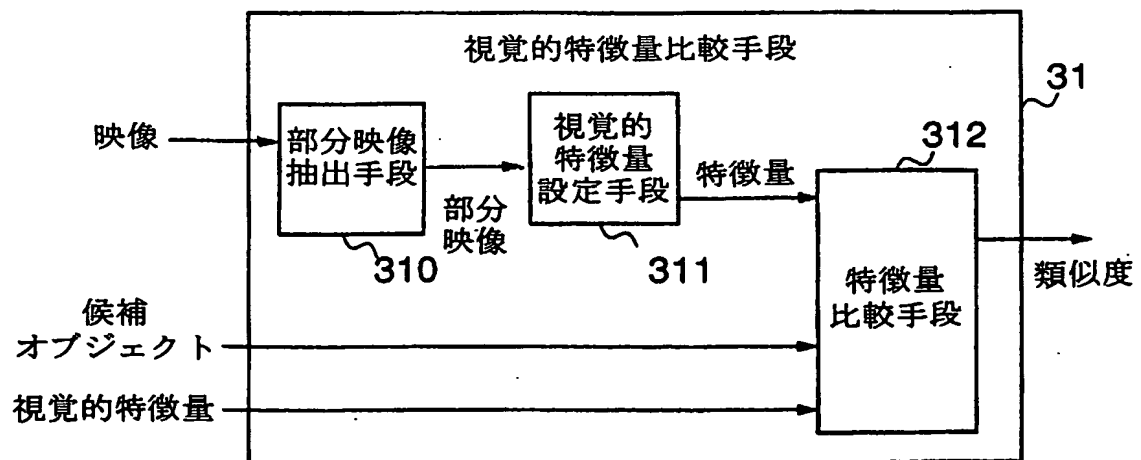


図 5

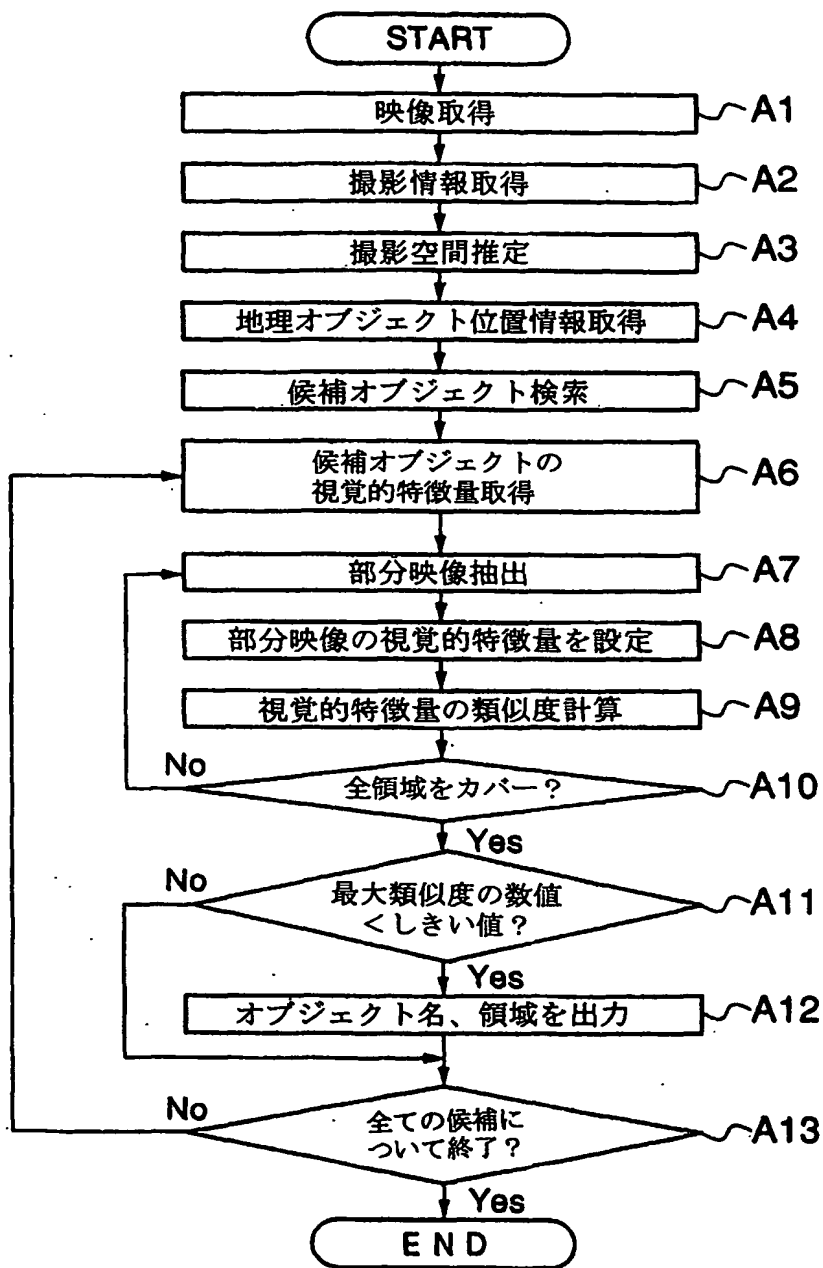


図 6

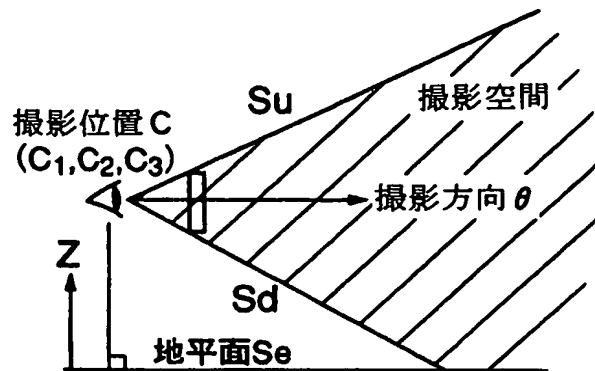


図 7

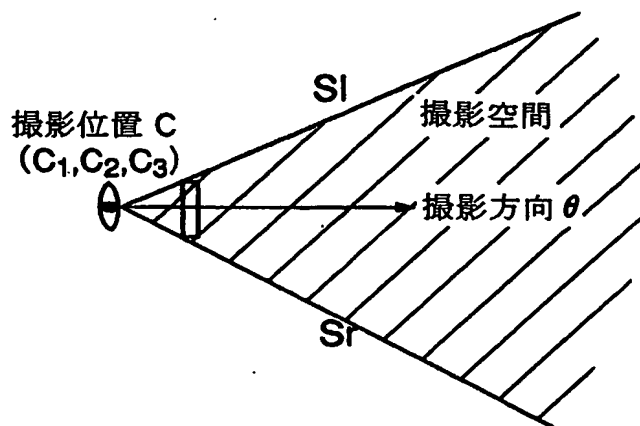


図 8

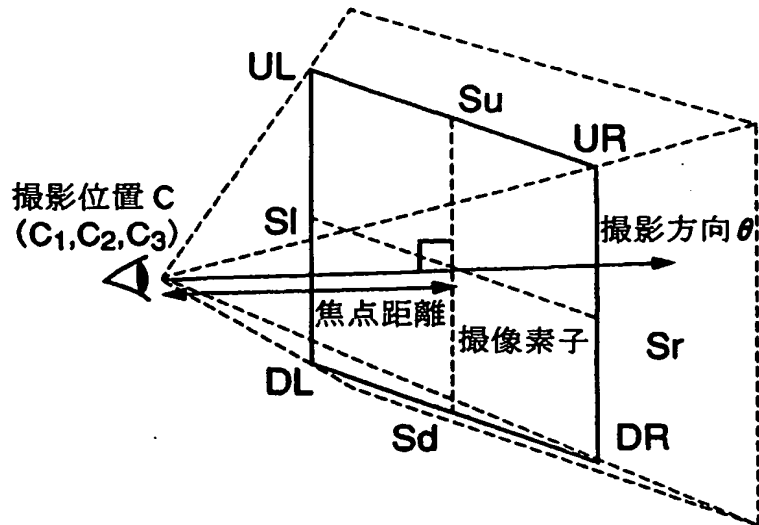


図 9

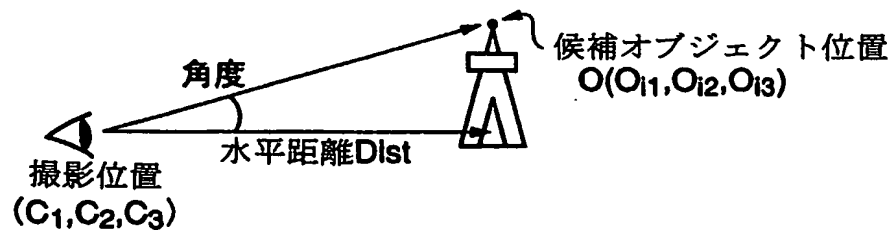


図 10

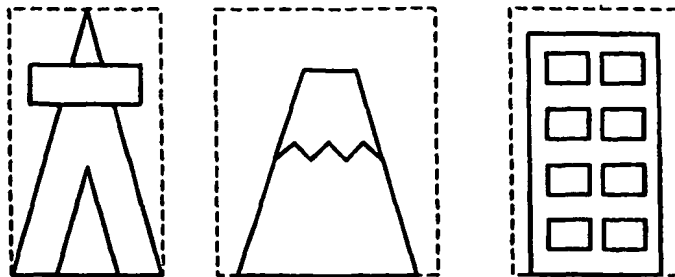


図 11

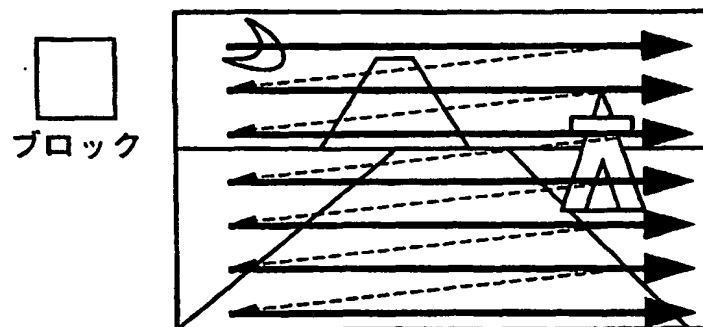


図 1 2

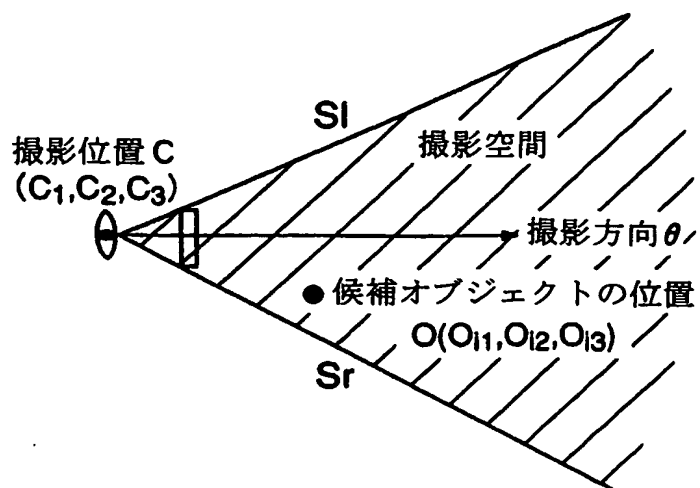


図 1 3

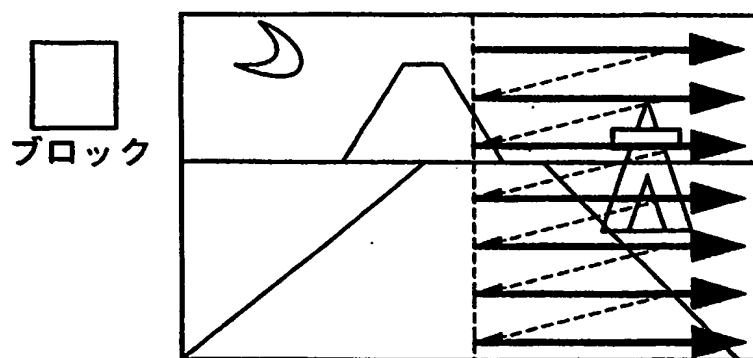


図 1 4

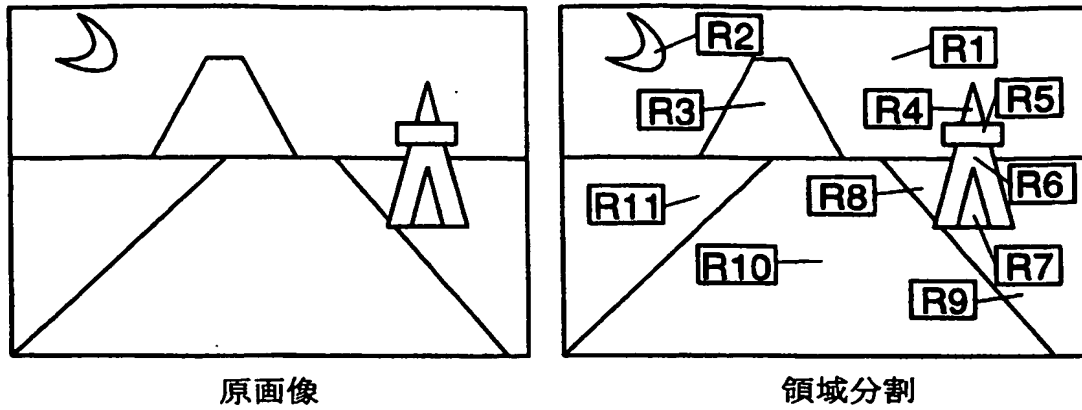


図 1 5



図 1 6

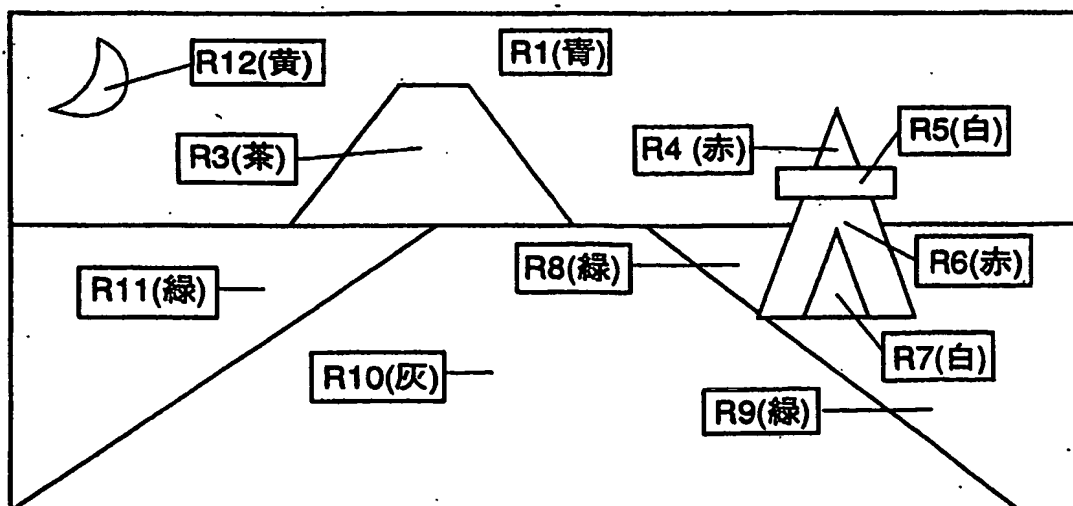


図 1 7

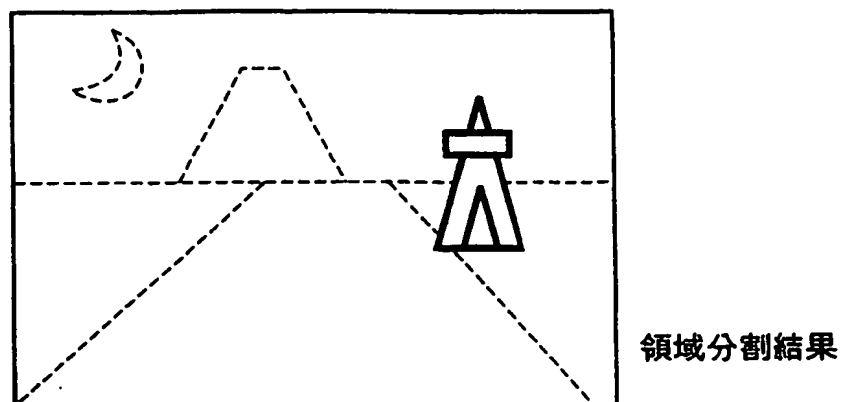


図 1 8

候補オブジェクト

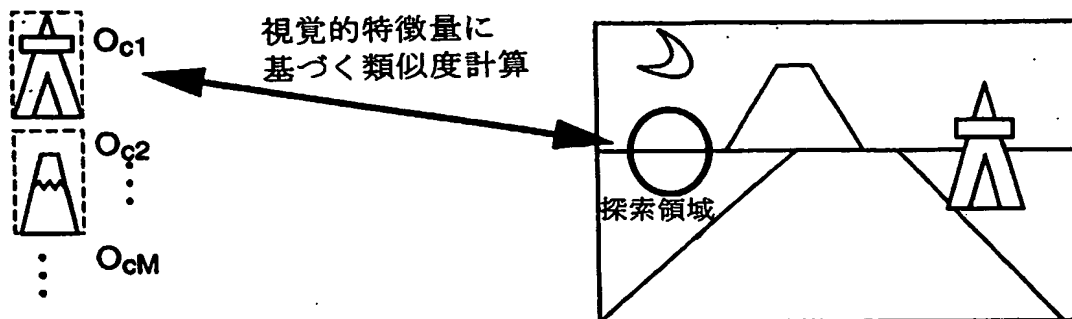
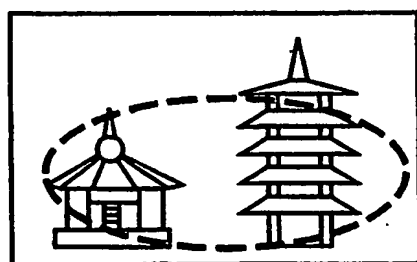
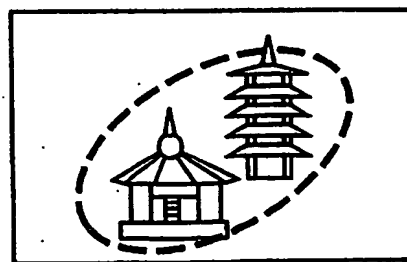


図 1 9



D 寺



D 寺



図 2 0

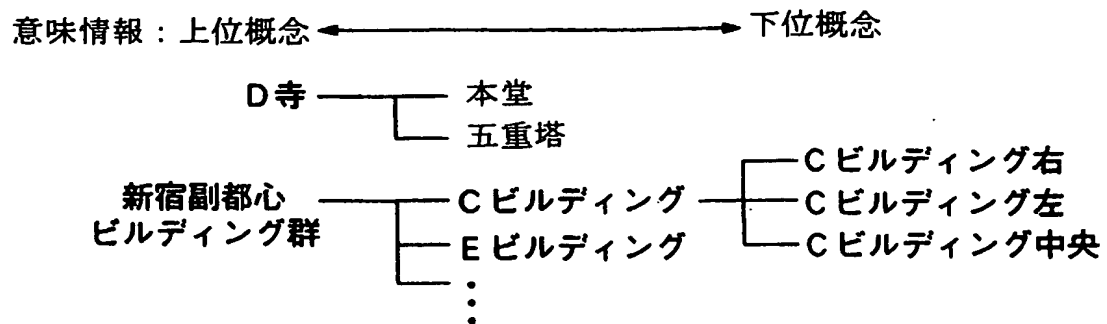


図 2 1

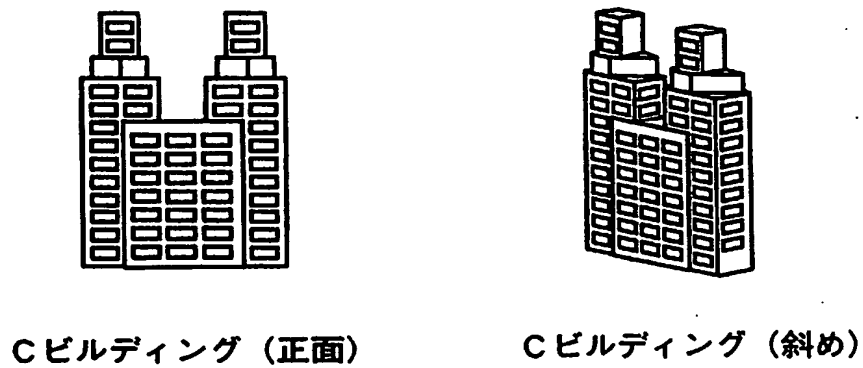


図 2 2

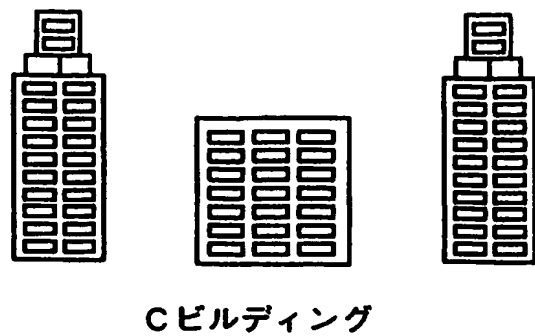


図 2 3

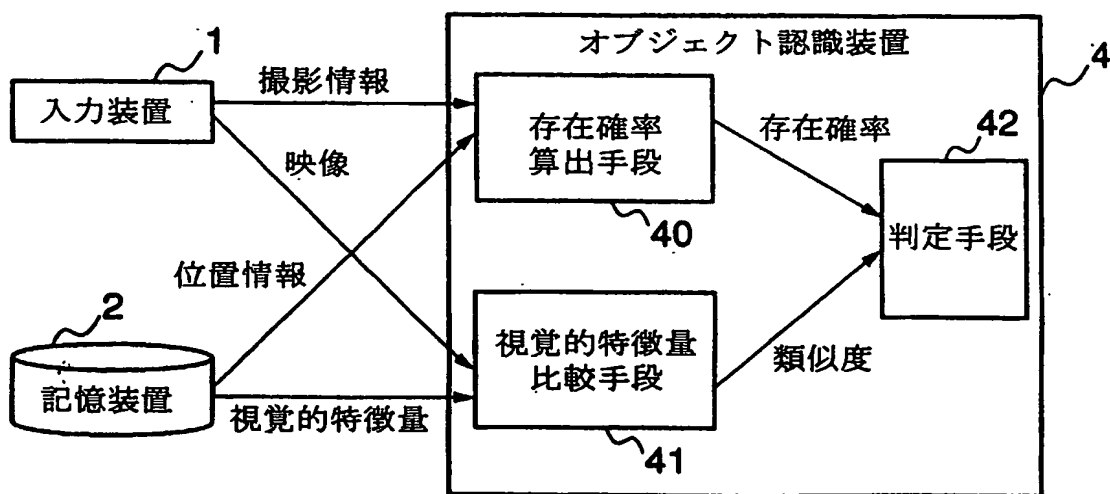


図 2 4

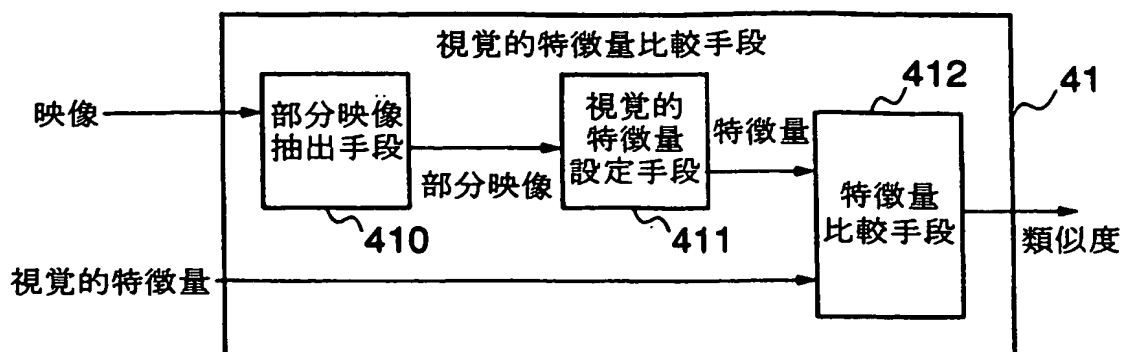


図 2 5

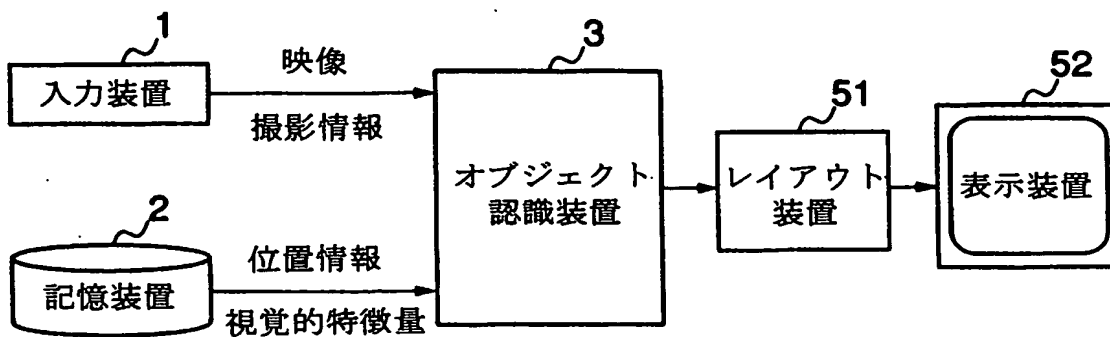


図 2 6

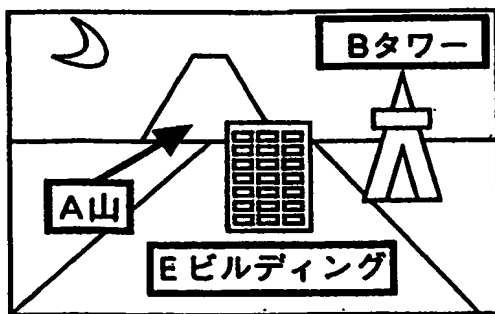


図 27

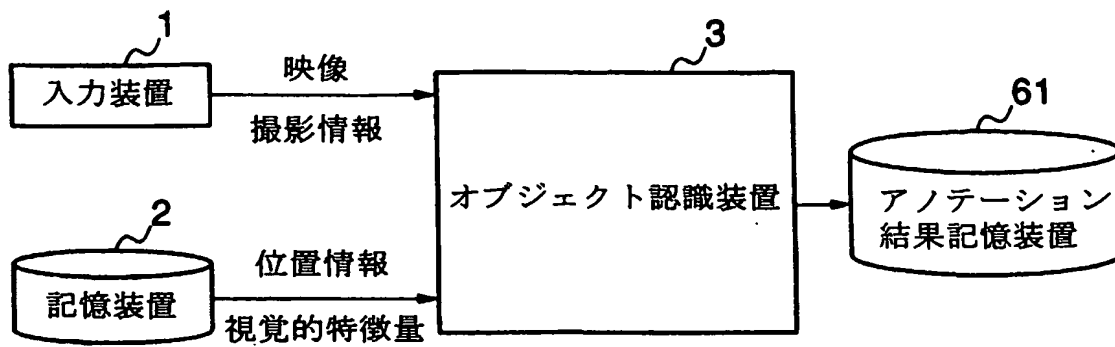


図 28

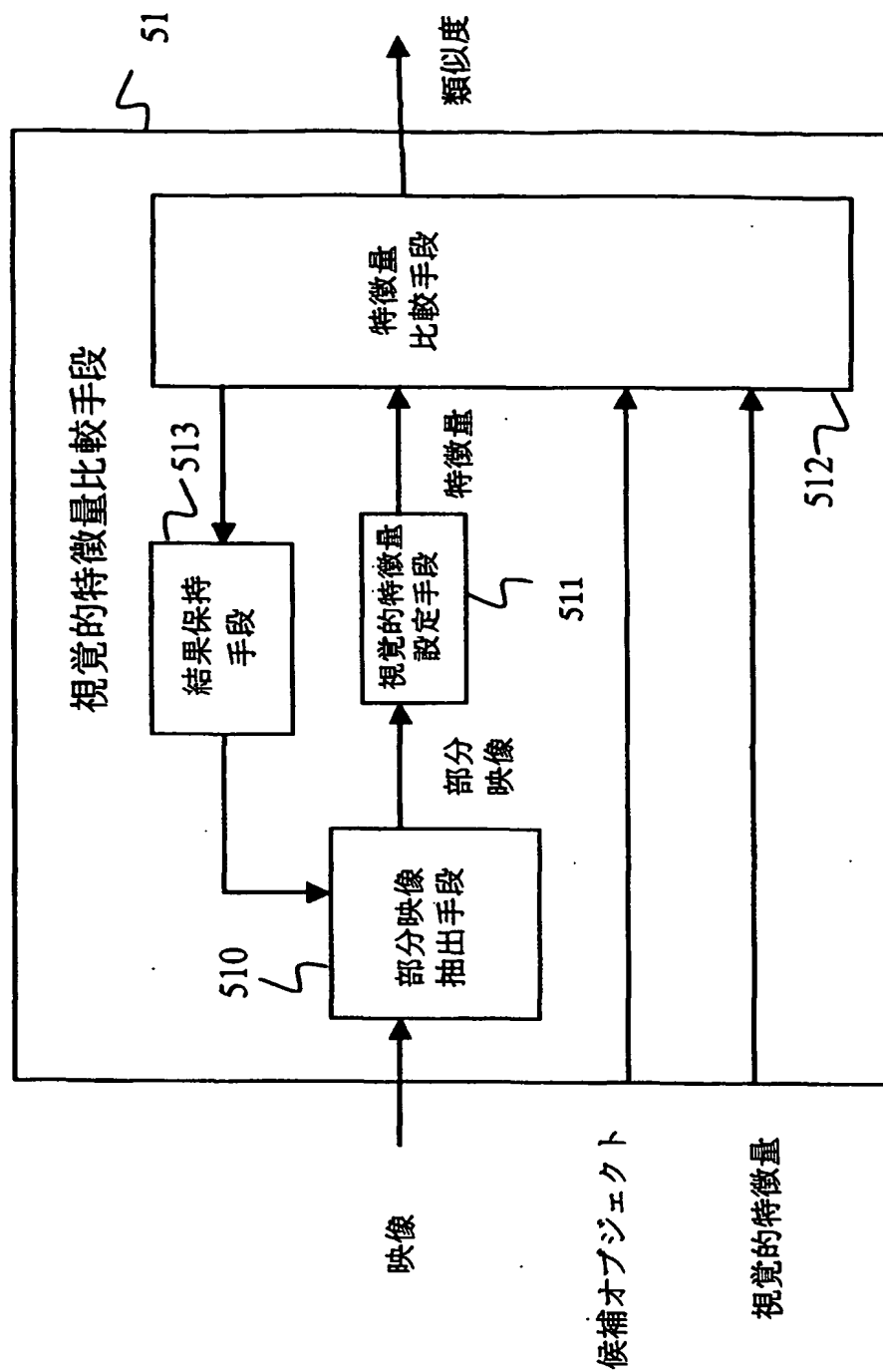


図 29

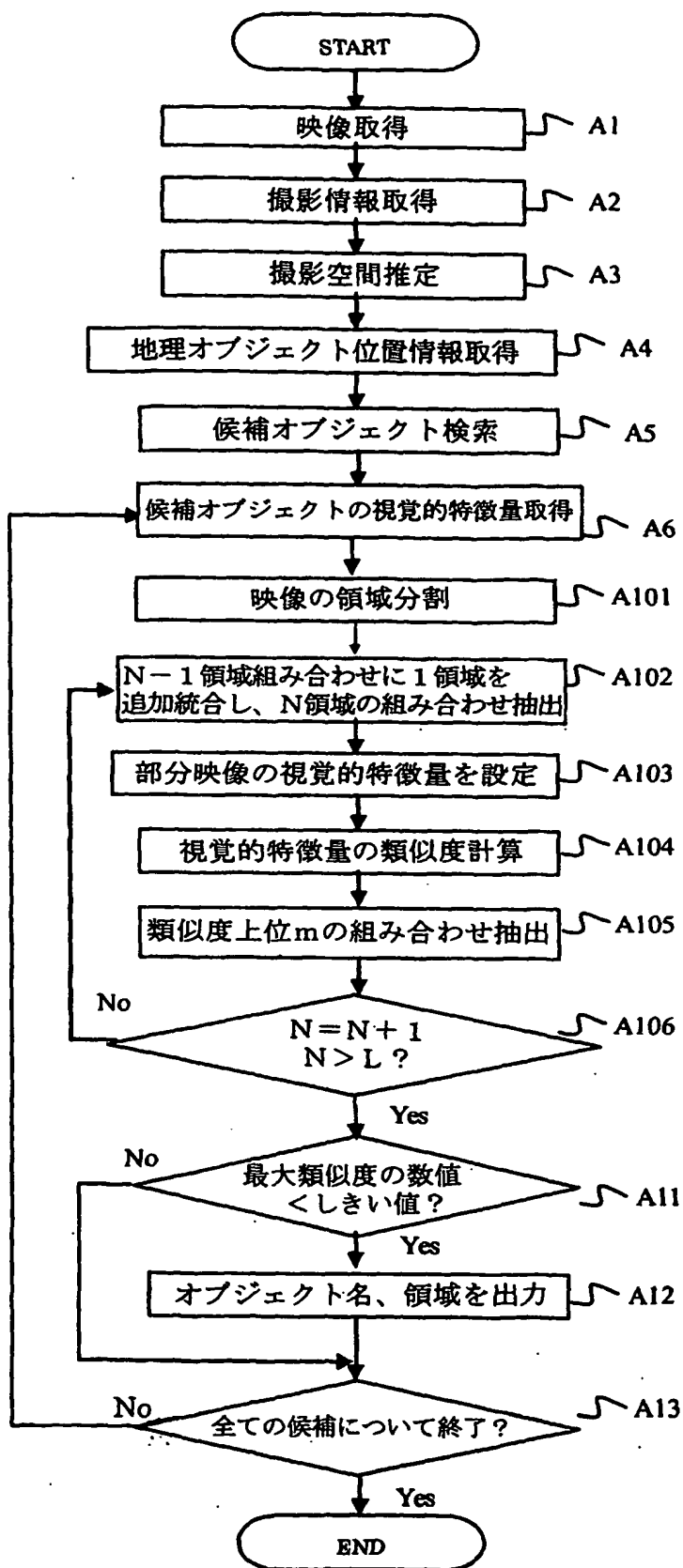
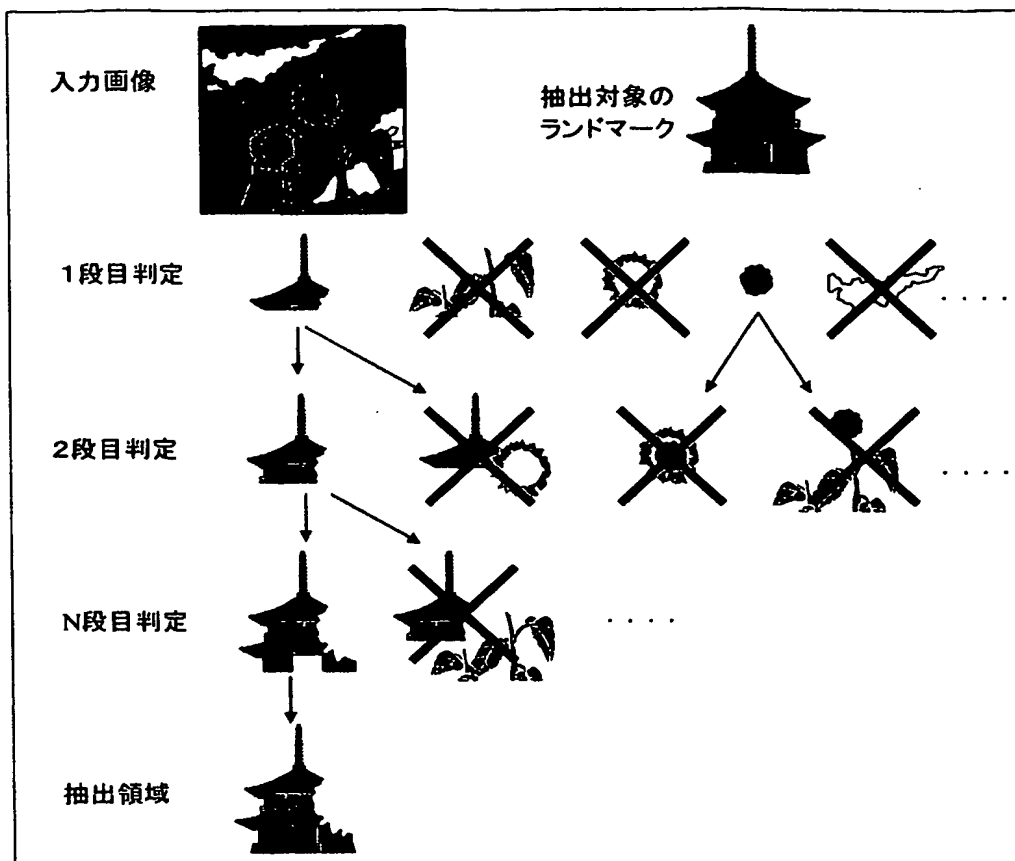


図 30



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/005372

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> G06T7/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G06T7/00-7/60

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2004
Kokai Jitsuyo Shinan Koho	1971-2004	Toroku Jitsuyo Shinan Koho	1994-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 10-267671 A (Nippon Telegraph And Telephone Corp.),	1, 2, 4-17,
Y	09 October, 1998 (09.10.98), Full text; all drawings & EP 1306648 A2 & US 6222583 B1 & CN 1203369 A	19, 20, 25-29 3, 18, 21-24
Y	JP 11-86035 A (Nippon Telegraph And Telephone Corp.), 30 March, 1999 (30.03.99), Full text; all drawings & EP 1306648 A2 & US 6222583 B1 & CN 1203369 A	3, 23

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search  
19 May, 2004 (19.05.04)Date of mailing of the international search report  
01 June, 2004 (01.06.04)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/005372

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2001-229382 A (Nippon Telegraph And Telephone Corp.), 24 August, 2001 (24.08.01), Par. No. [0030]; Fig. 3 (Family: none)	18
Y	JP 61-262986 A (Fujitsu Ltd.), 20 November, 1986 (20.11.86), Full text; all drawings (Family: none)	21-24

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G06T7/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G06T7/00 - 7/60

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2004年  
 日本国実用新案登録公報 1996-2004年  
 日本国登録実用新案公報 1994-2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 10-267671 A (日本電信電話株式会社) 1998. 10. 09, 全文 全図 & EP 1306648 A2 & US 6222583 B1 & CN 120336 9 A	1, 2, 4-17, 19, 20, 25-29
Y		3, 18, 21-24

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的な技術水準を示すもの  
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」 口頭による開示、使用、展示等に言及する文献  
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献。

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
 「&」 同一パテントファミリー文献

国際調査を完了した日

19. 05. 2004

国際調査報告の発送日

01. 6. 2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)  
 郵便番号 100-8915  
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

新井 則和

5H

8937

電話番号 03-3581-1101 内線 3531

## C (続き) . 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	JP 11-86035 A (日本電信電話株式会社) 1999. 03. 30, 全文 全図 & EP 1306648 A2 & US 6222583 B1 & CN 120336 9 A	3, 23
Y	JP 2001-229382 A (日本電信電話株式会社) 2001. 08. 24, 第0030段落 図3 (ファミリーなし)	18
Y	JP 61-262986 A (富士通株式会社) 1986. 11. 20, 全文 全図 (ファミリーなし)	21-24